The Dissertation Committee for Joshua David Slocum Certifies that this is the approved version of the following dissertation:

Measuring and Manipulating Electric Fields Near the Green Fluorescent Protein Fluorophore Using Vibrational Stark Effect Spectroscopy of Nitrile Probes

Committee:

Lauren Webb, Supervisor

Carlos Baiz

Andreas Matouschek

Pengyu Ren

Sean Roberts

Measuring and Manipulating Electric Fields Near the Green Fluorescent Protein Fluorophore Using Vibrational Stark Effect Spectroscopy of Nitrile Probes

by

Joshua David Slocum, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Acknowledgements

I would like to acknowledge the lab technicians in the molecular biology core facility for their excellent services in providing mass and sequence analysis on countless protein and DNA samples throughout my time in Austin. I would also like to thank my dissertation committee, Drs. Carlos Baiz, Andreas Matouschek, Pengyu Ren, and Sean Roberts, for their helpful discussions throughout the course of my graduate career. I would like to thank the entire Webb group for providing a rock-solid support team as I performed this research and for their valuable comments that ultimately helped shape the research into the form that is presented here. Specifically, I would like to acknowledge Elisa Novelli, Jeremy First, and Andrew Ritchie for numerous valuable discussions about the work herein, including Jeremy's vital efforts towards modelling many of the complex problems that are discussed below. I would like to thank my advisor, Lauren Webb, for her invaluable mentorship over the last five years. Lauren taught me how to think critically about *everything* and showed me how to effectively communicate my ideas to others. She allowed me the freedom to explore my own research interests, but was always an unwavering source of good ideas and motivation.

I thank my parents for their endless support, without which none of this would have been possible. And finally, I thank Heidi Culver for being a constant source of motivation to me as we have gone through graduate school together.

Measuring and Manipulating Electric Fields Near the Green Fluorescent Protein Fluorophore Using Vibrational Stark Effect Spectroscopy of Nitrile Probes

Joshua David Slocum, PhD

The University of Texas at Austin, 2017

Supervisor: Lauren Webb

The nitrile stretching oscillation has been widely used as a probe of local environment to study dynamics, folding, and electrostatics in proteins. A popular model for interpreting nitrile frequencies is the vibrational Stark effect (VSE), which allows one to interpret changes in vibrational energy in terms of changes in force along the nitrile bond. In principle, this allows for the site-specific, directional measurement of electric fields in a complex protein environment. However, the interpretation of these frequencies in terms of electric fields is complicated by the fact that hydrogen bonding to the nitrile probe is known to cause frequency shifts that are not described by the VSE. To address this concern, we have biosynthetically incorporated *para*-cyanophenylalanine (*p*CNF) probes into green fluorescent protein (GFP) near the intrinsic fluorophore, whose sensitivity to electric fields has been well characterized. We observed that the vibrational and electronic probes of electrostatic environment have similar spectroscopic responses to a series of amino acid mutations, and that the intrinsic sensitivity of GFP emission energy to the mutations was unperturbed by the presence of the *p*CNF probes.

Additionally, we compared the measured Stark effect shifts to pK_a changes of the GFP fluorophore and saw that these two orthogonal measurements of electrostatic environment were in agreement, which further corroborates the analysis of VSE shifts of nitriles in terms of electric fields. These studies provide confidence in the ability of nitrile frequencies to report faithfully on electric fields, even in the background of direct hydrogen bonding to the probes. Finally, based on our analysis of electrostatic forces near the GFP fluorophore, we designed and characterized an interesting GFP mutant that can be photoactivated by near-UV light. We obtained preliminary evidence in the form of mass spectrometry and kinetic analysis that support a novel charge transfer mechanism for this photoconversion, which could lead to the design of fluorescent proteins with enhanced properties. In summary, we have developed GFP as a robust model system for understanding and manipulating the finely tuned relationship between protein structure and function.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The complex arrangement of partially-charged atoms in a protein structure gives rise to a heterogeneous electrostatic environment. The precise direction and magnitude of these electrostatic forces are integral to every aspect of protein function, including folding, stability, enzymatic activity, dynamics, and multi-molecular interactions.[1–3] As such, there has been a great deal of effort directed towards the accurate measurement and prediction of the electrostatic properties of proteins, the complete understanding of which would allow for the design of better drugs, the treatment of many diseases, and the design of proteins with improved or new functions. High-resolution structural data for thousands of proteins has helped to elucidate the relationship between protein structure and function by revealing buried active sites of enzymes[4] and shape complementarity in protein-protein interfaces,[5] for example. However, structural data alone does not allow one to visualize the non-covalent interactions that mediate structure-function relationships. In fact, because of the spatial heterogeneity in the electrostatic environments of proteins, the accurate estimation of electrostatic properties is a challenge from both experimental and computational viewpoints. This dissertation provides a detailed report of the work we have done towards the measurement and manipulation of electric fields in model fluorescent proteins using a variety of spectroscopic techniques. This introductory chapter will provide the necessary background for understanding the context of the work presented here.

1.1: Measuring Electrostatics in Proteins Using pK_a Shifts of Titratable Residues

The most widely implemented experimental technique aimed at measuring electrostatic effects in proteins has been the measurement of the perturbation of pK_a values of titratable amino acid side chains compared to their values in aqueous buffer.[6–10] The three-dimensional arrangement and connectivity of partially charged atoms near the titratable residue creates an electrostatic environment that strongly influences the relative stability of charged *versus* neutral side chain. Because of this, the measurement of the equilibrium between these two states is often thought to be a good reporter of the average protein environment around the amino acid. pK_a values can be measured by recording the pH-dependence of a signal that changes with protonation state; this is commonly done by measuring chemical shifts of labile ^{1}H or nearby ^{13}C atoms by NMR.[10] Experiments like these have led to the estimation of dielectric constants of globular proteins,[11,12] helped rationalize the catalytic mechanisms of enzymes,[13–15] and have led to an understanding of how charged amino acids can be accommodated in the interior of proteins.[8,12,16]

While these pK_a values are excellent reporters of the overall electrostatic environment around the titratable residue, it is typically very difficult to rationalize which specific interactions contribute most significantly to the observed equilibrium. For instance, lysine, whose side chain pK_a is ~10.5 in water, often titrates with a significantly lower pK_a when buried in the non-polar interior of a protein. This is due to the overall energetic favorability of the non-polar protein environment to contain the neutral form of lysine rather than the charged form. However, without further experimental and theoretical investigation, the perturbed pK_a values of lysine residues do not offer

significant information about the molecular level detail of the specific electrostatic interactions that cause the pK_a shift.

Because of the sensitivity of pK_a values of titratable amino acid side chains to the overall electrostatic environment around the amino acid, one significant aspect of accurate pK_a measurements in proteins is their usefulness as benchmarks for electrostatics models.[17–24] However, this sensitivity can also be viewed as a drawback for modeling, because of the necessity to accurately model long range interactions in a complex protein environment. Many attempts have been made to predict the pK_a values of different amino acids in proteins to validate electrostatics models with levels of theory ranging from fully atomistic quantum mechanics, to mixed quantum mechanics, to continuum electrostatics.[17,18,25–27] However, because of the complexity of electric forces in proteins, the accurate modeling of electrostatic effects often requires excessive computational cost or a vast oversimplification of the relevant physics. Most electrostatics models can adequately reproduce the pK_a values of solvent exposed side chains that experience small shifts in pK_a relative to their values in water, but the successful calculation of perturbed pK_a values of buried residues, which have been observed to shift by as much as 5 pK_a units, is much more difficult.[25,28] This requires an accurate and detailed treatment of the complex environment of a protein interior, including dynamic fluctuations.[25,27,29–31]

Many current strategies for calculating pK_a shifts in proteins involve the use of a thermodynamic cycle to calculate the change in free energy when the titratable residue in question is transferred from water to the protein environment.[19,32,33] The strategy taken to accurately calculate these electrostatic free energies is where many common models

deviate, although most are based on continuum electrostatics models such as the Poisson-Boltzmann (PB) equation. Algorithms to generate distributions of charge states and starting structures for the energy calculations generally increase the agreement with experiment, which suggests that protein fluctuations and charge state coupling is important.[22,23] However, regardless of the approach, electrostatics models still do a poor job at reproducing pK_a shifts of buried residues. Indeed, in a recent collaboration between many researchers, blind predictions of the pK_a values of 25 lysine residues in the interior of *Staphylococcal* nuclease were made using many different strategies, each meeting only limited success.[27,30] Most of the techniques reported in the study were based on continuum electrostatics models that differed in their approach to sampling conformational and charge space of the protein atoms. While many of the approaches could accurately predict the direction of the pK_a shifts relative to the value in water, often the magnitude of the pK_a shifts were off by several pK_a units. Because many of the current strategies depend on an accurate calculation of the total electrostatic energy of the protein, which is sensitive to many factors such as protein conformation, dipole moments, and polarizability, it is often difficult to assess which aspects of a given model contribute to both the agreement and disagreement between experiment and calculation.

1.2: Measuring Electrostatic Fields in Proteins Using Stark Effect Spectroscopy

An increasingly popular method for measuring electrostatic effects in proteins is Stark effect spectroscopy, which is also known as the electrochromic effect and refers to the sensitivity of spectroscopic transitions to electric fields. The Stark effect is entirely

analogous to the Zeeman effect, which refers to the sensitivity of spectroscopic transitions to magnetic fields and is regularly exploited in NMR spectroscopy. In principle, the response of a probe transition to an electric field can be estimated by a Taylor series expansion about the applied field shown in Equation 1:

$$E = -\left(\Delta\vec{\mu} \cdot \vec{F} + \Delta\alpha \cdot \vec{F} \cdot \Delta\alpha + \cdots\right) \qquad \text{[Equation 1.1]}$$

where E is the absorption energy of the transition, $\Delta\vec{\mu}$ is the difference in the dipole moment of the probe associated with the transition, $\vec{F}$ is the applied electric field, and $\Delta\alpha$ is the change in polarizability associated with the transition. By recording the spectrum of a molecule under the application of a known value of applied electric field, the parameters in Equation 1 can be estimated.[34,35] Boxer and coworkers have characterized the values of $\Delta\vec{\mu}$ (also called the Stark tuning rate) and $\Delta\alpha$ for a wide range of probe transitions, including the chromophores in photosynthetic reaction centers[36,37] and fluorescent proteins,[38,39] Ruthenium dyes,[40,41] tryptophan,[42] heme,[43] nitric oxide,[44] carbonyls,[45,46] nitriles,[47–49] C-F bonds, and C-D bonds.[50]

For many of the investigated vibrational transitions, as well as the electronic transitions of some fluorescent proteins, the electric field response is dominated by the difference dipole term in Equation 1.[47,48] When this is the case, it may be possible to directly relate the energies of the probe transition to the strength of the electric field it experiences because the response is linear. However, because the probe transition energy is related to the applied field through a dot product, it is necessary to have structural data about the orientation of $\Delta\vec{\mu}$ (typically achieved through molecular dynamics simulations).

Typical values of $\Delta\vec{\mu}$ range from ~0.05-0.10 D for vibrational transitions to ~5-10 D for electronic transitions. A more convenient set of units to express these difference dipole moments in is cm^{-1}/(MV/cm), which can be arrived at by scaling the difference dipole moment by Planck's constant and the speed of light. In these units, typical vibrational Stark tuning rates are ~1 cm^{-1}/(MV/cm), which means that the vibrational absorption energy of a probe transition will shift by 1 cm^{-1} for every 1 MV/cm value of electric field experienced by the probe. The corresponding Stark tuning rates for electronic transitions are around 100-200 cm^{-1}/(MV/cm). These units are convenient, because simulations have shown that the heterogeneity of electric fields in proteins is on the order of tens of MV/cm.[51–53]

Once a probe transition's Stark tuning rate has been characterized, it is common to use changes in its absorption energy to gain information about the electric field due to a specific perturbation, relative to a well-defined reference. Thus, we can re-write Equation 1 in the form shown in Equation 2:

$$\Delta E = -\Delta\vec{\mu} \cdot \Delta\vec{F} \qquad \text{[Equation 1.2]}$$

where ΔE is the change in absorption energy of the probe transition and $\Delta\vec{F}$ is the change in electric field. The Stark effect is commonly used in this way to obtain relative electric fields in proteins due to protein-protein docking, amino acid mutation,[54] ligand binding,[49,55] and location-dependent changes as illustrated in Figure 1.1.[56,57]

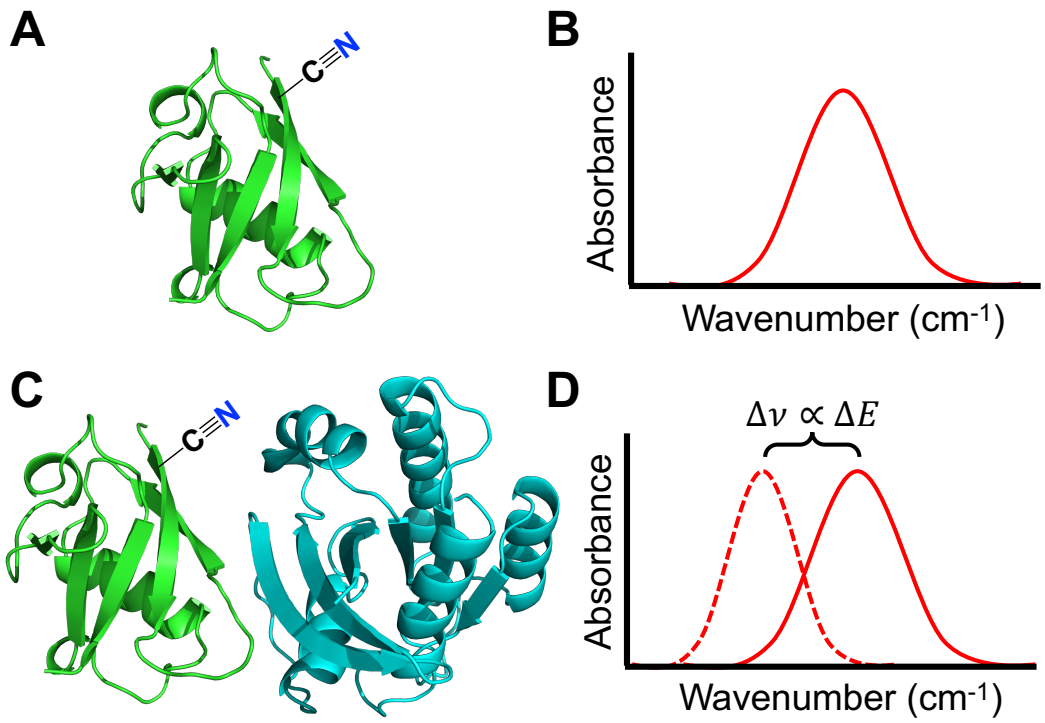

Figure 1.1: Cartoon representation of the use of Stark effect spectroscopy to measure the change in electric field induced by a protein-protein interaction. A monomeric protein that has been site-specifically labelled with a nitrile Stark effect probe is shown in (A). A simulated spectrum of the nitrile probe on the monomeric protein is shown in (B). Panel (C) shows the binding of a second protein at the nitrile-containing interface. Panel (D) shows a simulated spectrum of the nitrile probe in the docked complex. The change in energy of the Stark effect probe between states (A) and (C) can be used to estimate the change in electric field due to the protein-protein interaction based on Equation 1.2.

For these calibrated probe transitions to be useful for measuring electric fields in a protein environment, they must meet several criteria. First, the probe transition must be easily incorporated into the protein of interest. This requirement is easily met by the photosynthetic reaction centers, fluorescent protein chromophores, and heme transitions, as these probe transitions are native to their respective proteins. Additionally, carbonyl groups are present in the backbone of all protein chains. Nitrile groups, as well as C-F bonds, are common in naturally-occurring ligands or drug molecules that have been designed to interact with proteins. For studies where there is not an intrinsic probe transition whose electric field response has been well-characterized, there is an increasingly long list of methods for incorporating these functional groups into proteins in a site-specific manner.[58–61]

Once a probe transition has been incorporated into the protein of interest, the second criterion is that its spectrum be easily separated from the background of the protein spectrum. This means the absorption cross section should be strong and the energy of the transition should be well-removed from native protein transitions. For most of the electronic Stark effect probes, this is not a problem because they absorb near-UV or visible light (to which most proteins are transparent) with high extinction coefficients (ε ~10^4-10^5 M^{-1} cm^{-1}).[37,62] Nitrile, C-D, and azide vibrational absorptions have significantly lower extinction coefficients (ε ~10^2 M^{-1} cm^{-1}), but have absorption energies (~2000-2300 cm^{-1}) that are well-removed from the lower energy protein vibrations.[47,50] Together, this makes it relatively straightforward to measure vibrational signals from

these labels. In contrast, carbonyl probes must be isotope labelled in order to differentiate their absorption energies from the amide backbone vibrational energies.[46]

Two final considerations for these probes are their sensitivity to electric fields (i.e. Stark tuning rate) and their size. Because electronic transitions involve the movement of electrons, they are accompanied by very large changes in dipole moment. This is the reason for the two order of magnitude difference in measured Stark tuning rates of electronic transitions compared to those of vibrational transitions. However, labelling a protein with a bulky dye could be potentially drastically perturbative to native electrostatics and function. In contrast, the vibrational probes represent minimal perturbation, being composed of only two atoms. While the vibrational Stark tuning rates are much smaller, changes in vibrational energy of a few cm^{-1} can easily be measured. Overall, the Boxer group has characterized the electrostatic responses of a wide-ranging set of probe transitions that have been used to address a variety of questions related to protein electrostatics.

Site-specific measurements of electric fields in proteins using these calibrated probe transitions may provide a more suitable experimental handle on electrostatic effects in proteins than alternative measurements like pK_a or NMR shifts. This has largely to do with the fact that in most cases Stark effect shifts can be directly related to electric fields.[50,57,63] This is in contrast to pK_a or NMR shifts, which are only indirectly related to their non-covalent environment as described in the previous section. Moreover, an electric field is a much more concrete physical quantity than a 'non-covalent interaction'. A site-specific electric field quantifies the direction and magnitude of an electric force

that would be experienced by a positive test charge, which is a variable that can easily be calculated from computation models. For this reason, electric field measurements are an ideal benchmark for electrostatics models.

Perhaps the most well-known use of Stark spectroscopy in a protein is work by Steffen, et. al. on the dielectric environment of the photosynthetic reaction center.[37] Structural data at the time had shown that there were two possible electron transfer routes in the reaction center. However, electron transfer only occurs efficiently along one route. Stark spectroscopy of the symmetric chromophores in the photosynthetic reaction center revealed that the dielectric environments between the two were significantly different, such that electron transfer would only happen in one direction. This observation would not have been possible with structural data alone. More recently, Boxer and coworkers have measured electric fields in the active site of keto-steroid isomerase using a variety of carbonyl and nitrile vibrational probes.[57,63] Estimates of active site electric fields have been compared to calculated fields in order to provide benchmarks for computational models. Additionally, for the first time, an estimate of the contribution of active site electric fields to the rate of a catalytic reaction has been made.

1.3: Vibrational Stark Effects of Nitriles

Nitrile VSE probes are of particular interest because of their relatively large Stark tuning rates (0.4-0.8 $cm^{-1}/(MV/cm)$), large absorptivities ($\varepsilon \sim 500\ M^{-1}\ cm^{-1}$), and absorption in a region of the infrared well removed from a protein's background vibrations ($\upsilon_{CN} \sim 2100\text{-}2240\ cm^{-1}$).[47,50] To facilitate these experiments, a number of

methods for incorporating nitriles into proteins or peptides have been developed, such as chemical synthesis, posttranslational modification, or nonsense suppression, and a considerable effort from the Boxer group has led to the measurement of $\Delta\vec{\mu}$ for a wide range of different nitriles.[47,49,50] Additionally, because $\Delta\vec{\mu}$ for the nitrile stretch is parallel to the nitrile bond vector, the measured change in absorption energy can be directly related to the change in field in the direction of the nitrile bond.[48]

Many groups have made use of the nitrile vibration to measure electrostatic effects in many different protein systems. One of the earliest efforts, by Getahun, et. al., made use of a *p*-cyanophenylalanine-labeled calmodulin-binding peptide.[33] By comparing the spectra of the nitrile-labeled peptide before and after binding to calmodulin, it was determined that the nitrile in the free peptide was well-solvated by water and then buried in a hydrophobic pocket upon binding. This was one of the first demonstrations that nitrile labels could be used to gain information about their local environments in biological systems. Suydam, et. al. reported the first quantification of electric fields at the active site of an enzyme by measuring the vibrational spectra of nitrile-containing ligands that bound to the active site of human aldose reductase.[51] Additionally, they compared the VSE measurements to electric fields calculated from molecular dynamics simulations and showed that agreement between the measured and calculated electric fields could be achieved only after the inclusion of extensive dynamics and ensemble averaging. More recently, efforts by Boxer and coworkers have focused on the measurement of electric fields at the active site of ketosteroid isomerase by measuring VSE shifts of nitriles introduced with a variety of techniques.[56,57]

Much of the interest in our group has been on the measurement of electric fields at protein-protein interfaces. Stafford, et. al. introduced thiocyanate-based probes onto the surface of the Ral guanine dissociation stimulator (RalGDS) and measured VSE shifts of the nitriles due to the binding of two different GTPases.[64] They found that VSE shifts of some of the nitrile probes were similar between the two GTPases, while some were very different, which suggested that key residues at the interface are involved in an electrostatic mechanism for differential binding of RalGDS to either of the GTPases. Walker, et. al. performed a similar study with thiocyanate probes on the surface of a GTPase to investigate the electrostatic effects of a known structural tilt in the binding orientation of the GTPase to two different downstream effectors.[65] Many of the datasets of electric fields at protein surfaces and protein-protein interfaces generated in these studies have served as benchmarks for computational studies aimed at accurately calculating protein electric fields. Ritchie and Webb have shown that continuum electrostatics models can be used to recreate differences in electric field, but the accurate calculation of absolute fields that are in agreement with nitrile frequencies has so far remained a challenge.[66,67]

Despite the extraordinary progress that has been made in the development of nitrile probes for measuring protein electric fields, their use is somewhat diminished by their ability to accept hydrogen bonds, which, while being electrostatic in origin, causes deviation from the Stark effect model due to quantum effects associated with the hydrogen bond. Indeed, it has been shown in non-hydrogen bonding solvents that the nitrile frequencies of acetonitrile and benzonitrile are inversely proportional to solvent

dielectric.[68,69] However, in water and other protic solvents, the frequencies are not well correlated with solvent dielectric or any other solvent property. [63,70] *Ab initio* calculations of acetonitrile and methyl thiocyanate in water have also revealed that nitrile frequencies can span as much as 25 cm^{-1} depending on the geometry of the hydrogen bond between nitrile and water.[71] This sensitivity to hydrogen bonding, which arises from changes in the force constant of the nitrile bond upon hydrogen bonding rather than changes in the difference dipole moment, casts doubt on the efficacy of nitriles as VSE probes of electric field. This is a major concern when attempting to compare experimentally measured frequencies to electric fields calculated from simulations,[66,67,72] and several attempts have been made to deconvolute these competing effects. Fafarman, et. al., have shown that by comparing nitrile frequencies to chemical shifts of isotope labeled nitriles, it is possible to subtract the hydrogen bonding contribution to the frequency and determine the portion of a frequency shift that is only due to electrostatic field effects.[73] However, in spite of this advancement in the interpretation of nitrile VSE experiments, there is still not agreement on the ability of nitrile frequencies to report directly on electric fields. It has been suggested that nitrile stretches in biological systems are predominantly sensitive to their local hydration status and are most useful for determining if a particular position is buried or solvent exposed.[72,74]

Additionally, the ability of nitriles to accept hydrogen bonds leads to the possibility of introducing hydrogen bonds where there previously were none, which could grossly misrepresent the behavior of the native system, the understanding of which is the ultimate goal. Adhikary, et. al., have shown that nitrile probes can be particularly de-

stabilizing, with buried probes causing more destabilization than solvent-exposed ones.[72] This measure of protein destabilization due to a nitrile probe represents an attempt to quantify the amount of perturbation due to the probe, which is an effect that needs to be investigated whenever a non-native probe is used to interrogate a system. For proteins it is common to assay the function with and without the probe, and as long as the function is not altered significantly it is assumed that the probe does not significantly perturb the system under investigation.[54,57,75] However, measurements like these only indirectly assess the extent of perturbation caused by the probe. If the insertion of a nitrile probe drastically perturbs the electrostatic environment of the protein, for example by the addition of its large ground state dipole moment, then electric field measurements based on the nitrile probe frequencies are not likely to lead to any understanding of the electrostatics in the native protein. Additionally, it is possible that an electrostatic probe could cause local perturbations in electrostatic field without affecting the observed structure, stability, or function, and so an assay of any of these properties would not reveal problematic changes caused by the nitrile itself. In the case of a site-specific nitrile VSE probe, it would be desirable to know what the electrostatic field is in the absence of the probe.

In this dissertation, we address these concerns using green fluorescent protein (GFP) as a model system. Specifically, we have been interested in validating the response of nitrile VSE probes by comparing frequency shifts to alternative measurements of electrostatic environment, especially when the nitrile probes are engaged in hydrogen bonds. GFP has given us an excellent model system for doing this.

1.4: Green Fluorescent Protein as a Model System for Measuring Electrostatic Effects

Green fluorescent protein (GFP), shown in Figure 1.2, is one of the most widely-studied proteins due to its unique fluorophore, which forms upon folding in the presence of oxygen.[76,77] Initially isolated from *Aequoria victoria* jellyfish in the Pacific Northwest, GFP has gained widespread use due to the ability of the fluorophore to form without the need for any jellyfish-specific co-factors. This means the GFP gene can be expressed and result in the production of a fluorescent protein in almost any organism, making it extremely useful as a genetic imaging tag.[78] The wild type (WT) fluorophore, shown in Figure 1.2C, is shielded from bulk water by 11 β-strands and forms from the amino acids at positions 65-67: S65, Y66, and G67 (where S is serine, Y is tyrosine, and G is glycine; these one-letter amino acid codes will be used throughout the rest of the text). The absorption spectrum of WT GFP has two major features: an intense, high energy band around 395 nm and a weaker, low energy band around 488 nm, called the A and B state, respectively. Crystallographic and spectroscopic data have shown that the A and B state absorption peaks correspond to Y66 being either protonated or de-protonated, respectively.[77,79] Absorption of either the A or B state leads to efficient fluorescence around 510 nm. However, the A state does not fluoresce directly. Rather, it undergoes excited state proton transfer before fluorescing.[38,77,80] The overall photocycle of GFP is shown in Figure 1.3.

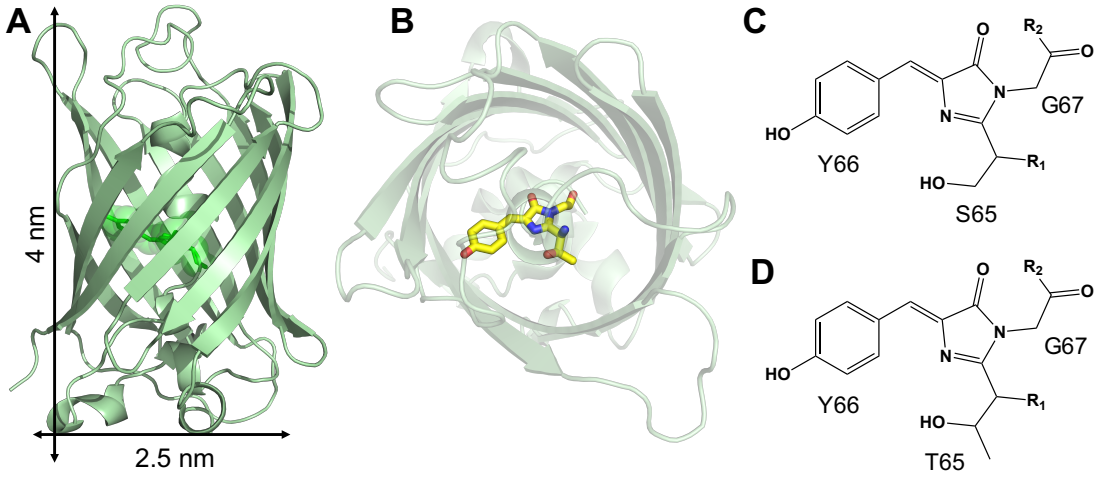

Figure 1.2: The structure of green fluorescent protein (GFP) and its embedded fluorophore. (A) The crystal structure of superfolder GFP (2B3P) with the backbond rendered as a cartoon. 11 beta strands make the iconic beta barrel structure. (B) A top-down view of GFP showing the interior of the barrel and the embedded fluorophore. (C) The serine-tyrosine-glycine fluorophore. (D) The threonine-tyrosine-glycine fluorophore.

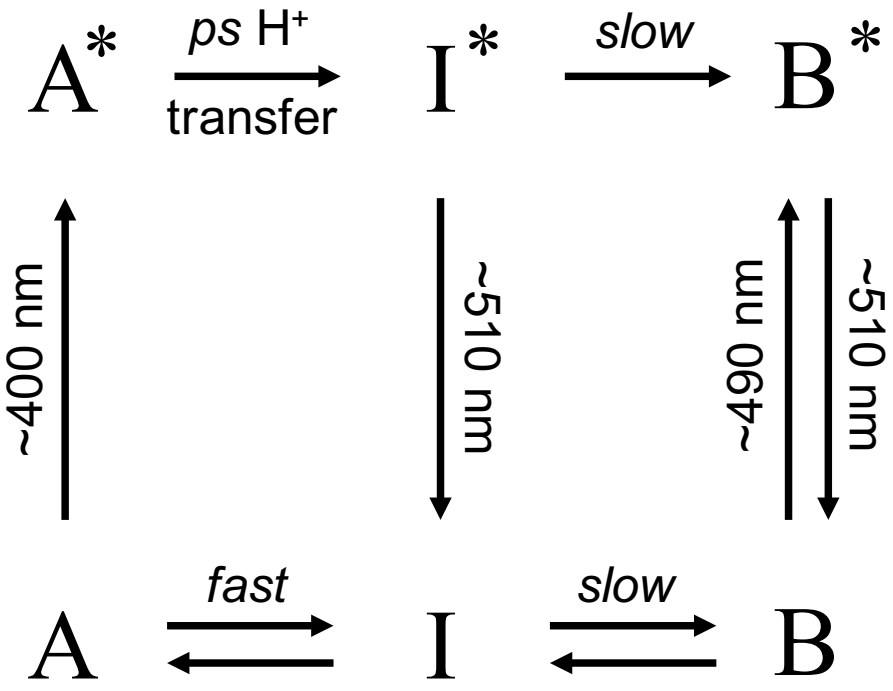

Figure 1.3: Photocycle of canonical GFP mutants. In the ground state, conversion between the A and B states proceeds through an intermediate called the I state. A and I interconvert quickly relative to the conversion between I and B, because the B state has a significantly rearranged hydrogen bonding network involving many of the nearby residues. The excited A state usually does not fluoresce directly, rather it rapidly undergoes excited state proton transfer to the excited I state before fluorecsing. The excited I state fluoresces at a similar wavelength to that of the excited B state.

Significant directed evolution and engineering efforts have led to fluorescent protein (FP) sequences that are related to WT GFP, but that have significantly altered photophysical and biochemical properties.[81–83] Several key residues in the amino acid sequence have emerged from these studies as being integral to the fluorophore photophysics. Of particular interest to us here are the residues at positions 65, 203, and 222.[84,85] They will be described below as they are relevant to the work presented. However, the interested reader is referred to several reviews which provide more depth on the FP sequences that have been developed, including those isolated from different species.[80,86–89]

The variant of GFP that we have chosen to work with here is called superfolder GFP (sfGFP), which gets its name from its engineered ability to fold quickly and remain stable when fused to poorly folding proteins.[90] Additionally, sfGFP has a high thermal stability and a high tolerance for random amino acid mutations. sfGFP contains one particular mutation, S65T (which denotes serine at position 65 being replaced with threonine), that causes significant changes in its spectral properties relative to WT GFP. Specifically, the S65T mutation causes the fluorophore to mature faster and for the B state absorption to dominate over the A state absorption. High resolution crystal structures have shown that in GFP mutants containing S65, the A state is stabilized by a hydrogen bonding network involving a water molecule and S205, and the negative charge on E222.[77] However, in the mutants containing T65, the extra methyl group on the fluorophore causes a steric hindrance that stabilizes a hydrogen bonding network

ultimately inhibiting the ionization of E222.[91] Thus, in the T65 mutants, E222 is neutral and the fluorophore exists predominately in the B state.

The ultimate utility of the TYG fluorophore for our purposes is that the B state transition was shown to be dominated by the linear term in Equations 1 and 2, with a Stark tuning rate of 117.5 $cm^{-1}/(MV/cm)$.[39] Thus, the intrinsic fluorophore of sfGFP constitutes an electronic Stark effect probe that we can use as a probe of electric fields. Additionally, the side chain of T203 is optimally oriented such that its replacement with an aromatic residue forms a stabilizing interaction with the fluorophore that shifts the emission to lower energy. Indeed, it has been shown that incorporating different residues at position 203 can shift the emission by as much as 20 nm.[92] In this work, we commonly use position T203 substitutions to create large changes to the electrostatic environment near the GFP fluorophore.

One final feature of GFP that we will investigate in this work is the light driven decarboxylation of E222. It was noted very early on that exposure of GFP to near-UV light caused an increase in the B state absorption at the expense of the A state.[78,85] Crystallographic and spectroscopic studies have revealed a mechanism for this that is initiated by electron transfer (ET) from E222 to the excited state of the fluorophore.[93–95] The fluorophore then donates an electron and a proton back to E222, which results in the loss of CO_2. The overall effect of this is a change in fluorophore equilibrium due to the neutralization of E222. Patterson and Lippincott-Schwartz screened for mutants that displayed maximum contrast before and after exposure to near-UV light and arrived at a mutant containing S65 and H203, which they called photoactivatable GFP (PA-GFP).[82,96]

The photoactivation property of PA-GFP was then used in one of the first demonstrations of superresolution imaging in live cells by Betzig, et. al.[97] Since then, engineering efforts have led to the development of many different FPs which can be photoactivated following this same mechanism, with analogous electron transfer between the embedded fluorophore and a nearby glutamic acid residue.[88,98–100] In Chapter 5, we will report several variants of sfGFP which display photoactivation with a similar efficiency as PA-GFP but that contain T65-based fluorophores. Our work suggests that photoactivation properties in FPs can be further enhanced by considering electron transfer from sites other than E222.

1.5: Outline of the Dissertation

The following chapters in this dissertation will report the work that we have done using GFP as a model system for measuring electrostatic fields in proteins using biosynthetically-incorporated nitrile probes. An overall theme is to address the aforementioned concerns with using nitriles as electric field probes. Specifically, we are interested in hydrogen bonding to the nitrile probes, and relating the nitrile probe responses to other types of electrostatic probes, such as the GFP fluorophore absorption energies and its pK_a values. Chapter 2 thoroughly outlines the procedure for expressing and purifying *p*CNF-labeled GFP, as well as the experimental procedures for all of the data generated herein. In Chapter 3, the results of a set of electric field measurements near the GFP fluorophore are reported. We compare the changes in electric field, caused by a series of amino acid mutations, measured from *p*CNF probes and the GFP

fluorophore. The results show that the two types of field probe respond similarly to the electrostatic perturbations, even when the nitrile probes appear to engage in hydrogen bonds. Chapter 4 details the comparison of nitrile VSE shifts in GFP to measurements of the pK_a of the GFP fluorophore. We show that these two distinct types of electrostatic probe have well-correlated responses to a series of nearby mutations and discuss cases where the responses are not correlated. In Chapter 5, we build on our understanding of GFP fluorophore equilibrium to generate and characterize several mutants of GFP which display interesting photoactivation properties. We discuss these findings in the context of an electron transfer mechanism that is common to fluorescent proteins across many species. Finally, in Chapter 6 we report several interesting observations that show promise as potential future research directions.

Chapter 2: Experimental Methods

2.1: Publication Note

Portions of the methods outlined in this chapter were adapted from the following publications:

1. J. D. Slocum, J. T. First, L. J. Webb, "Orthogonal Electric Field Measurements near the Green Fluorescent Protein Fluorophore through Stark Effect Spectroscopy and pK_a Shifts Provide a Unique Benchmark for Electrostatics Models." *J. Phys. Chem. B, 121 (28)*, 6799-6812 (2017). [J. D. Slocum performed all experimental measurements, including protein expression and purification, infrared and UV-Vis spectroscopic measurements, and pK_a measurements].
2. J. D. Slocum and L. J. Webb, "A Double Decarboxylation in Superfolder Green Fluorescent Protein Leads to High Contrast Photoactivation". *J. Phys. Chem. Lett., 8 (13),* 2862-2868 (2017).
3. J. D. Slocum and L. J. Webb, "Nitrile Probes of Electric Field Agree with Independently Measured Fields in Green Fluorescent Protein Even in the Presence of Hydrogen Bonding". *J. Am. Chem. Soc., 138 (20),* 6561-6570 (2016).

2.2: Introduction

In this chapter, the experimental methods of protein expression and purification, as well as all spectroscopic measurements, are outlined. One key difference in the VSE experiments presented here, compared to previous experiments in the Webb group, is the method of nitrile probe incorporation. In these experiments, we have made use of amber suppression technology to incorporate *para*-cyanophenylalanine (*p*CNF) into GFP. This does not add many experimental difficulties besides the need to grow *E. Coli*. cells in minimal media and to incorporate two plasmids into the cells: one that contains a

precisely placed amber codon, and one that contains the machinery to suppress the amber codon. Overall, amber suppression technology has emerged as a powerful tool for unnatural amino incorporation. Several recent reviews provide a more thorough background for the interested reader.[101–103]

Amber suppression, which is sometimes called nonsense suppression, simply refers to the suppression of the amber codon.[58] The amber codon is one of three codons that do not code for any amino acid. In this regard, it is a nonsense message, because it does not code for any letters of the genetic alphabet. Rather, the amber codon is a stop codon, which means that when it is read by the ribosome during the translation of RNA into protein, translation stops. Therefore, amber suppression means to allow the amber carry to carry a message other than 'stop'. In order to do this, orthogonal tRNA-synthetase pairs are generated.[104,105] The tRNA is the biomolecule which delivers a particular amino acid to the ribosome in response to a particular codon. The synthetase is the enzyme which catalyzes the attachment of a designated amino acid to a designated tRNA. The design of an amber suppression system is therefore straightforwardly based on the generation of an orthogonal pair that contains a tRNA which can recognize the amber codon, and a synthetase which can attach an unnatural amino acid (in our case *p*CNF) to the tRNA.[104]

With this machinery in place, it is possible to genetically encode the new unnatural amino acid. We were given two plasmids by Ryan Mehl (Oregon State), who has developed many of the methods that this Chapter is based on. One plasmid contains the gene for superfolder GFP (sfGFP), and one which contains the gene for an orthogonal tRNA-synthetase pair optimized for *p*CNF.[59] These two plasmids were expressed simultaneously in *E. Coli*. cells to incorporate *p*CNF in place of the amber codon on the sfGFP gene. By changing the location of the amber codon on the sfGFP gene, we were

therefore able to genetically encode for the incorporation of *p*CNF into sfGFP in a site-specific manner.

The amber suppression technology has been around for quite some time now and is rapidly changing. Many of the recent improvements have been manifested in better incorporation yields, which are ultimately very important for infrared spectroscopy experiments with weak absorbing oscillators like the nitrile stretch. Initially, we attempted to incorporate *p*CNF into a Ras GTPase. However, the incorporation efficiency was not high enough to yield a significant amount of nitrile-labelled protein for FTIR spectroscopy experiments. In contrast, the methods below are routinely employed to prepare *p*CNF-labelled sfGFP with yields on the order of ~500 mg/L of growth media.

2.3: Site Directed Mutagenesis and Plasmid Purification

The gene for His_6-tagged sfGFP on a pBAD plasmid was generously provided by Ryan Mehl (Oregon State). In order to express GFP with a *p*CNF probe, mutations were made to the pBAD gene to place the *amber* stop codon (TAG) at the desired position. Any other mutations were made ensuring that the E. Coli optimized codon was used. Mutations were made using a Quick-Change site-directed mutagenesis kit from Agilent and short primers containing the desired mutation, purchased from Sigma. Varying concentrations of template DNA (between 10-50 ng/μL) were mixed with 125 ng of the primers, 5 μL of supplied buffer, 5 μL of dimethyl sulfoxide, 100 ng of free nucleotides, and water to a total volume of 50 μL. The reaction mixtures were centrifuged at 13000 RPM for 30 seconds, 1 μL of the supplied polymerase was added, and the mixtures were centrifuged for 30 more seconds at 13000 RPM. The reaction mixtures were then cycled 18 times according the following scheme: 95 °C for 30 seconds, 55 °C for 60 seconds, 68

°C for 360 seconds, and 55 °C for 30 seconds. After the 18th cycle, the reactions were held at 68 °C for 600 seconds and then were stored at 4 °C until digestion.

Digestion of the reaction mixtures was carried out by adding 1 μL of Dpn1 restriction enzyme from the mutagenesis kit, centrifuging for 30 seconds at 13000 RPM, and then incubating at 37 °C for one hour. After incubation, 10 μL of each of the digested reactions was added to 40 μL of XL1-Blue competent cells, and the mixtures were allowed to rest on ice for 30 minutes. The remaining digested reactions that were not added to cells were stored at -20 °C, to be used at a later time if needed. After the 30-minute rest period, the XL1-Blue cells were heat shocked at 42 °C for exactly 45 seconds and then were allowed to rest on ice for 2 minutes, after which 250 μL of 37 °C SOC media was added to each mixture. The cells were then grown for 1 hour at 37 °C, while shaking at 250 RPM, after which the mixtures were added individually to agar plates containing LB growth media and ampicillin at 100 μg/mL. The cell-containing plates were then incubated overnight (typically 12-16 hours) at 37 °C, after which they were wrapped in parafilm and stored at 4 °C until further use.

To generate enough DNA for sequencing, a single colony from each of the plates was used to seed 5 mL of LB growth media containing ampicillin at 100 μg/mL. The 5 mL cultures were allowed to grow overnight (typically 12-16 hours) at 37 °C and shaking at 250 RPM, after which the cells were pelleted by centrifugation at 3000 RPM for 15 minutes. The pBAD plasmid was purified from the resulting cell pellet using a Mini-Prep kit from Qiagen exactly as described in the manual. The concentration of purified DNA was estimated using an ATR-UV/Vis NanoDrop, and the sequences were verified by submitting 11 μL of the purified DNA and 1 μL of a T7 sequencing primer to the ICMB sequencing facility.

2.4: Cell Transformation

Genes that contained no TAG codon, and thus would not express *p*CNF-containing sfGFP variants were transformed into either BL21-DE3 competent cells or DH10β competent cells. In either case, 1.25 μL of purified DNA was added to 50 μL of cells, and the mixture was allowed to rest on ice for 30 minutes. After the rest period, the cells were heat shocked at 42 °C for 45 seconds before being allowed to rest on ice for 2 minutes. 250 μL of 37 °C SOC media was then added to the cells, which were then grown for 1 hour at 37 °C while shaking at 250 RPM. The cells were then plated onto warmed (37 °C) agar plates containing LB media and ampicillin at 100 μg/mL by dividing the cell mixture into equal volumes and plating onto two separate plates. The plates were incubated at 37 °C for 12-16 hours, and were then wrapped in parafilm and stored at 4 °C until further use (no longer than 3 weeks).

Genes that contained the TAG codon, and thus would express *p*CNF-containing mutants, were co-transformed into DH10β cells with a pDULE plasmid (provided by Ryan Mehl) that codes for the *p*CNF-tRNA/*p*CNF-tRNA synthetase pair. 1.25 μL of each plasmid was added to 50 μL of DH10β cells, which were then allowed to rest on ice for 30 minutes. The cells were then heat shocked at 42 °C for 45 seconds, before being allowed to rest on ice for 2 minutes. 250 μL of 37 °C SOC media was then added to the cells, which were then grown for 1 hour at 37 °C while shaking at 250 RPM. The cells were then plated onto warmed (37 °C) agar plates containing LB media and 100 μg/mL ampicillin and 25 μg/mL tetracycline by dividing the cell mixture into equal volumes and plating onto two separate plates. The plates were incubated at 37 °C for 12-16 hours, and were then wrapped in parafilm and stored at 4 °C until further use (no longer than 3 weeks).

2.5: Protein Expression

Single colonies on the plates from the previous section were used to seed growth in two types of media. E. Coli containing genes for GFP without a nitrile probe were grown in 1 L of TB media containing Ampicillin at 100 μg/mL and 0.5 g of arabinose. Cells that were co-transformed with the pDULE and pBAD plasmids, and would thus go on to generate nitrile-labeled proteins, were grown in 1 L of minimal media prepared following the recipe shown in Table 2.1. After 30 minutes of growth, the minimal media was supplemented with approximately10 mL of a 15 mg/mL solution of *p*CNF. 100 μL of 10M NaOH was added to the *p*CNF solution to help with dissolving. The composition of the 50x salt mixture and the trace metal stock solution are reported elsewhere.[60]

Autoclave the following in a 2800 mL Erlenmeyer flask
5 g of glycerol
1.2 mL of 40% (v/v) glucose
2 mL of 1 M $MgSO_4$
50 mL of 5% (v/v) aspartic acid (pH 7.5)
20 mL of 4 mg/mL leucine (pH 7.5)
880 mL of 18 MΩ water
After cooling, add the following
20 mL of 50x salt mixture
40 mL of 18 amino acid mixture
200 µL trace metal solution

Table 2.1: Minimal media recipe for expression of *p*CNF-containing mutants.

2.6: Protein Purification

The frozen cell pellets were thawed on ice and subjected to three rounds of sonication cycles that consisted of a one second pulse followed by a three second pause, for a total of four minutes. In between cycles, the cells were allowed to rest on ice for at least five minutes. The lysed cells were centrifuged at 19000 RPM for 30 minutes to remove the cell debris, and the resulting supernatant was passed through a lysate filter with a 10 μm pore size. The filtered lysate was added to a Ni-NTA affinity column that had been equilibrated with at least 12 column volumes of equilibration buffer containing 100 mM Hepes, 50 mM NaCl, and 30 mM imidazole at pH 7.5. After the total volume of lysate was added to the column, it was washed with 12 more column volumes of the equilibration buffer. His_6-tagged GFP was eluted from the column with 15 mL of elution buffer containing 100 mM Hepes, 50 mM NaCl, and 300 mM imidazole at pH 7.5. The eluted protein was then buffer exchanged into the equilibration buffer using a PD-10 desalting column. Cleavage of the His_6 affinity tag was performed by adding 0.25 mg/mL of trypsin to the protein solution and allowing it to incubate at 37 °C for 30 minutes, after which 0.5 mM of phenyl methyl sulfonyl fluoride (dissolved in isopropanol) was added to inhibit the protease reaction. The protein/affinity tag mixture was then added to a Ni-NTA column equilibrated in at least 12 column volumes of the equilibration buffer, and the cleaved GFP was collected directly from the flowthrough. Purified GFP was buffer exchanged into PBS at pH 7.5 (unless otherwise stated) and was either used immediately, or flash frozen and stored at -80 °C until further use.

2.7: Steady-State Absorption and Fluorescence Spectroscopy

UV-Vis absorption scans were performed using a Cary 5000 spectrometer and a 1 cm quartz cuvette. Circular dichroism spectra were collected on a Jasco J-815 spectrometer using a 1mm quartz cuvette and averaging over 10 scans. All FTIR spectra

were collected using a Bruker Vertex 70. The sample was injected into a cell between two sapphire windows separated by 125 μm Teflon spacers, and 250 scans were collected with a resolution of 0.5 cm^{-1}. Temperature dependent FTIR spectra were collected in a similar cell with 100 μm Teflon spacers. 200 scans were collected for *p*-tolunitrile in solvent and 600 scans were collected to achieve an adequate signal-to-noise ratio for the *p*CNF-containing proteins. Spectra were baseline-corrected using an in-house fitting program described previously.[75] *p*-tolunitrile and all solvents were purchased from Sigma and used without further purification. Fluorescence spectra were collected from samples in 1 cm polystyrene cuvettes with a Fluorolog3 spectrometer using a 1.5 nm slit width and a spectral resolution of 0.1 nm. All absorption and fluorescence measurements were averaged over at least three individual measurements.

2.8: pH-Dependent UV-Vis Spectroscopy

Purified GFP mutants were buffer exchanged into a master buffer containing 50 mM phosphate, 50 mM citrate, and 100 mM NaCl at pH 7.5, and the buffer-exchanged protein was concentrated to ~1 mM using centrifugal filters. UV-Vis absorption spectra of the GFP mutants were recorded from pH 5 to 10 on a Biotek Epoch 2 microplate reader between 250-600 nm. The plates were prepared by mixing 2 μL of the concentrated protein with 300 μL of the master buffer (adjusted to the desired pH with acid or base) into the wells of a UV-transparent 96-well plate. For each mutant, at least three pH titrations were carried out.

2.9: Time Correlated Single Photon Counting Measurements

Time-correlated single photon counting (TCSPC) was carried out using a Fluorolog3 spectrometer coupled with pulsed LED excitation sources. Samples were placed in 1 cm polystyrene cuvettes and were excited using LEDs at either 482 nm or 404

nm (unless otherwise specified). Emission was monitored using a monochromator with 0.5 nm slit width, and decays were collected until a minimum of 5000 counts was recorded for a single time point. The resulting decay data was fit using DAS6 Analysis from FluorEssence for either one or two exponential functions.

2.10: Photoactivation Experiments

Samples of purified GFP were photoactivated by exposure to a Spectroline XX-15G lamp for various amounts of time. Average power densities of the lamp, centered around 254 nm, of 58.7, 24.8, 6.9, 1.9, and 0.6 mW/cm^2 were achieved at distances of 4.1, 6.4, 12.1, 22.1, and 42.2 cm, respectively. Typical experiments were performed with ~100 μM of protein in PBS in a quartz cuvette, subjected to 10-60 second exposures for a total of 15-20 minutes. All photoactivation experiments were performed in at least triplicate on at least three independent protein expressions.

Chapter 3: Comparing Nitrile Probes of Electric Field to Independently Measured Fields in GFP

3.1: Publication Note

Portions of this chapter were adapted from:

J. D. Slocum and L. J. Webb, "Nitrile Probes of Electric Field Agree with Independently Measured Fields in Green Fluorescent Protein Even in the Presence of Hydrogen Bonding". *J. Am. Chem. Soc., 138 (20),* 6561-6570 (2016).

3.2: Introduction

There is a growing interest in using the nitrile vibrational oscillation as a site-specific probe of local environment to study dynamics, folding, and electrostatics in biological molecules such as proteins. Nitrile probes have been used extensively as reporters of electric field using vibrational Stark effect spectroscopy. However, the analysis of frequencies in terms of electric fields is potentially complicated by the large ground state dipole moment of the nitrile, which may irrevocably perturb the protein under investigation, and the ability of nitriles to accept hydrogen bonds, which causes frequency shifts that are not described by the Stark effect. The consequences of this are that vibrational spectroscopy of nitriles in biomolecules could be predominately sensitive to their local hydration status, not electrostatic environment, and have the potential to be particularly destabilizing to the protein. Here, we introduce green fluorescent protein (GFP) as a model system for addressing these concerns using biosynthetically incorporated *p*-cyanophenylalanine (pCNF) residues in the interior of GFP and measuring absorption energies of both the intrinsic GFP fluorophore and *p*CNF residues in response to a series of amino acid mutations. By making use of a system that contains both a vibrational (*p*CNF) and an electronic (GFP fluorophore) Stark effect probe, we investigated whether these probes gave a similar electrostatic response to the same set of

perturbations. Additionally, GFP's intrinsic linear sensitivity to electric field makes it useful as a model system for addressing the aforementioned challenges of using nitriles as Stark effect probes. Using nonsense suppression,[59,106] we inserted *p*CNF residues at positions 145 and 165 near the GFP fluorophore (Figure 3.1) and measured changes in absorption energy of these vibrational and electronic fluorophores in response to a series of mutations to position 203. We found that the change in fluorescence wavelength of GFP is strongly correlated to electric fields measured from both the vibrational and electronic Stark effect probes, and that similar types of mutations project similar changes in field onto all of the fluorophores. In addition, we used the GFP fluorophore to measure the extent of electrostatic perturbation caused by the *p*CNF probes. We observed that nitriles at these two locations can introduce changes in electrostatic environment around the TYG fluorophore, but that the changes are systematic and do not affect the fluorophores's intrinsic sensitivity to mutations at position 203. We then show that changes in electric field experienced by both of the *p*CNF probes are strongly correlated with changes in electric field measured by the fluorophore. Additionally, temperature dependent FTIR spectra of the nitrile-containing proteins suggest that the *p*CNF residues participate in hydrogen bonds even when buried in the interior of the protein. The observation that these nitrile probes give the same response as an independent field reporter to a series of mutations while engaging in hydrogen bonding suggests that Stark effect spectra of similar states may still be interpreted in electrostatic terms. While GFP is unique because of its intrinsic fluorophore, this work serves as an important example of how VSE of nitrile probes may be applied to other proteins of interest. This work was published in April 2016 and can be found at DOI: 10.1021/jacs.6b02156.

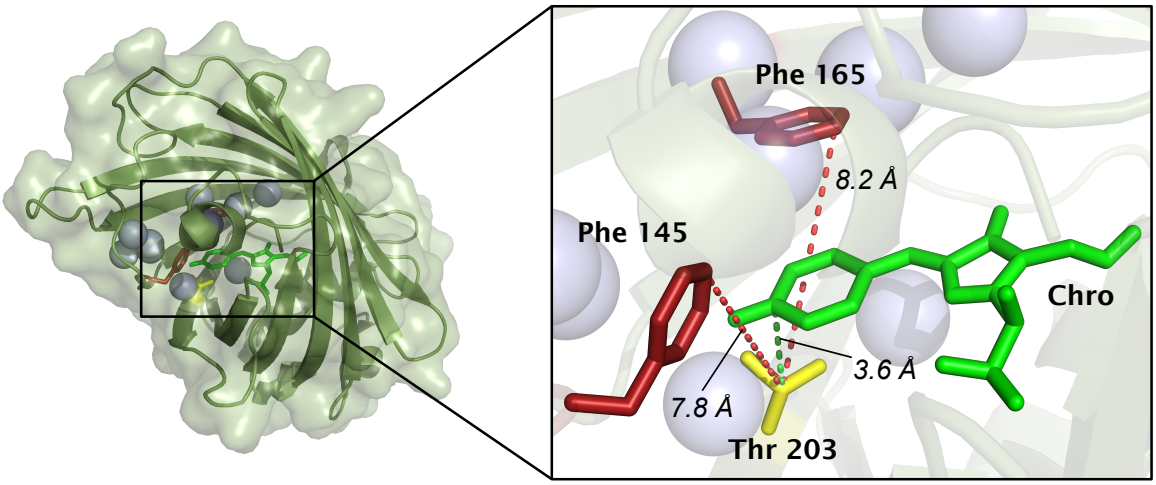

Figure 3.1: The crystal structure of superfolder GFP (PDBID: 2B3P). The backbone is shown as a ribbon structure and the light green is a representation of the solvent accessible surface area. Phe 145 and 165 are shown in red, Thr 203 is shown in yellow, and the fluorophore is shown in green. Eight ordered waters that are < 7Å from either residue 145 or 165 are shown as light blue spheres.

3.3: Vibrational and Electronic Absorption Spectra

We measured frequency changes of *p*CNF residues at two interior locations of GFP (positions 145 and 165), as well as frequency changes in both the absorption and emission of the deprotonated fluorophore, in response to five amino acid substitutions at position 203. Because the emission wavelength of GFP is known to be especially sensitive to mutations of amino acid 203, we reasoned that T203X mutations (where T is the original amino acid at position 203 and X represents the new amino acid: S, N, H, F, or Y) would create large changes in electric field that could be measured by both the electronic and vibrational chromophores. We chose positions 145 and 165 to insert the *p*CNF probes because the native phenylalanine side chains are optimally oriented on either side of the interaction between residue 203 and the fluorophore, and molecular modeling showed that nitriles at the *para* positions of the phenyl rings would point towards the fluorophore. Figure 3.2 shows crystal structure 2B3P with *p*CNF residues modeled into positions 145 and 165. The arrows represent $\Delta\vec{\mu}$ vectors, which point from nitrogen to carbon on the nitrile probes and from the hydroxyl group to the imidazolidone ring of the fluorophore. Additionally, Phe residues at these two positions are oriented nearly identically across several crystal structures containing the above T203X mutations, which suggests that *p*CNF residues inserted there might have a small range of motion.[79,107] As the Stark effect model allows us to interpret frequency changes as changes in field projected along a probe's difference dipole moment, it is desirable that the probe be as stationary as possible because the field projection depends on the probe's orientation in the field.

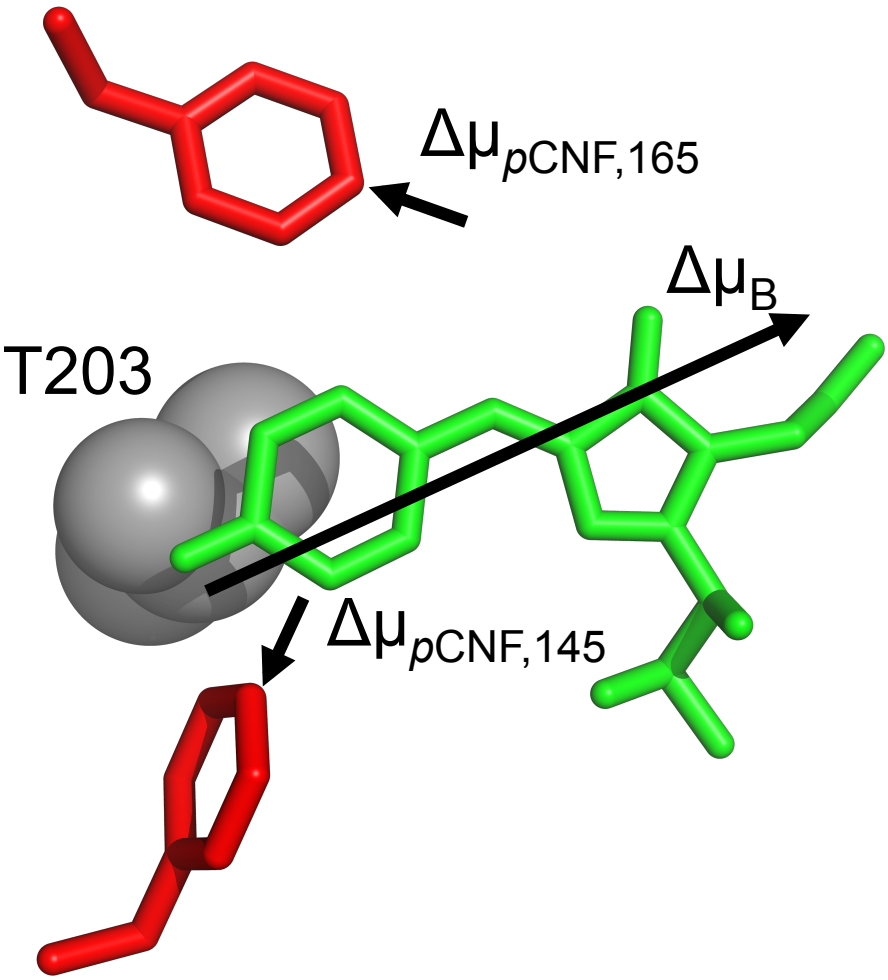

Figure 3.2: Close up view of crystal structure 2B3P highlighting Phe 145 and 165 (red), the GFP fluorophore (green), and Thr 203 (gray). Arrows represent the approximate magnitudes and directions of difference dipole moments of *p*CNF residues inserted at position 145 or 165, as well as the B state of the GFP fluorophore.

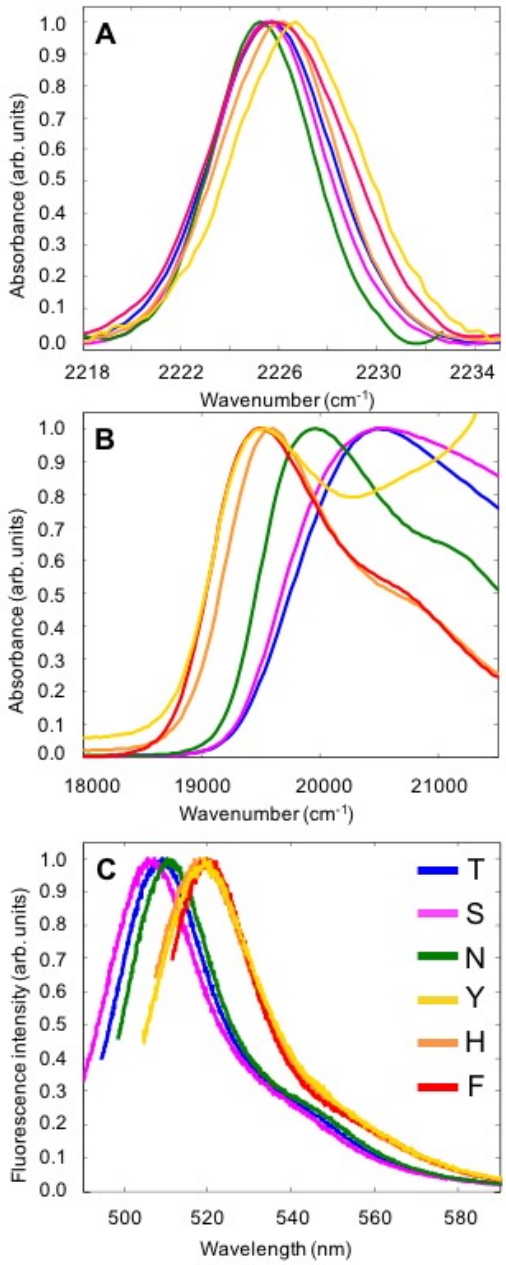

Figure 3.3: Representative spectra of nitrile-containing GFP constructs. A: Nitrile 145 absorption; B: fluorophore absorption; and C: fluorophore emission. Colors represent a different amino acid at position 203 based on the one letter amino acid codes shown inside panel (C).

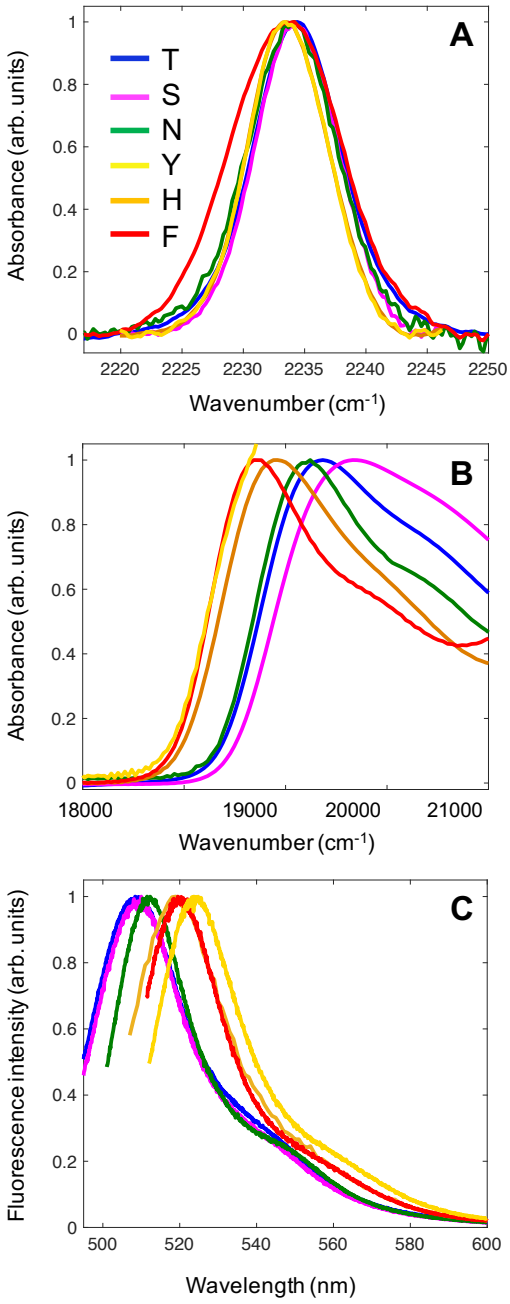

Figure 3.4: Representative spectra of nitrile-containing GFP constructs. (A) Nitrile 165 absorption; (B) fluorophore absorption; and (C) fluorophore emission. Colors represent a different amino acid at position 203 based on the one letter amino acid codes shown inside panel (C).

Representative spectra for the absorption of *p*CNF 145 GFP mutants, as well as absorption and emission of the intrinsic fluorophore are shown in Figure 3.3, and representative spectra for *p*CNF 165 mutants are shown in 3.4. As seen in Figure 3.3A, the FTIR spectra of the position 145 nitrile were well-described by a single Gaussian function with a FWHM of roughly 5-6 cm^{-1} (with the exception of the T203F mutant), which is similar to that for the free amino acid in a low dielectric solvent like THF.[70] Along with published crystallographic evidence, the narrow width of these bands suggests that each *p*CNF residue is likely confined to a region of space where it only sees a single, well-defined environment. For each nitrile position, the maximum shifts observed were on the order of 2 cm^{-1}. The absorption bands of the deprotonated fluorophore (Figure 3.3B) also exhibited a strong dependence on the T203X mutations, with maximum shifts on the order of ~1000 cm^{-1}. There was also more variability in the width of these bands. We observed narrower absorption peaks when the bulkier, aromatic residues (F, Y, and H) were at position 203, while the peaks were generally broadened when the smaller, polar side chains (T and S) were at position 203. This likely reflects a change in the pKa of the fluorophore's tyrosyl oxygen, which causes a change in the relative populations of the A and B states and results in a broadened B state absorption. Shifts in pKa such as these have been used previously to investigate electrostatic environments, and in future studies could prove to be a useful comparison to the frequency shifts measured here. Finally, representative emission spectra of the GFP variants are shown in Figure 3.3C, demonstrating that all mutations at T203X displayed the same major peak between 505-525 nm with a broader, less intense shoulder to the red.

3.4: Linking Vibrational and Electronic Stark Effects Via GFP Fluorescence

The usefulness of GFP for these studies lies in the intrinsic fluorophore, whose B state absorption energies have been shown to trend linearly with electric field strength. This intrinsic electronic Stark effect probe has a measured Stark tuning rate of 117.5 cm^{-1}/(MV/cm) oriented 21° away from the transition dipole moment, shown in Figure 3.2.[39,108] The Stark tuning rate of this electronic transition is much larger than that of the vibrational stretching transition of *p*CNF, which has a measured Stark tuning rate of 0.67 cm^{-1}/(MV/cm).[55] In addition, because the emission wavelength is dependent on the complex non-covalent interactions around the fluorophore, it serves as a convenient probe to assess the effect of inserted *p*CNF probes on these interactions. We therefore wanted to use the emission wavelength of GFP to check the electrostatic responses of *p*CNF probes against those of the electronic fluorophore.

To measure the effect on electric field due to T203X mutations, we have compared the changes in emission energy of the fluorophore between all pairs of mutants against the projection of changes in field onto either the vibrational or electronic probe ($\Delta\vec{F}$ from Equation 1.2, where the local field factor is assumed to be 1). Here, the six different T203X mutants give rise to 15 individual differences in energy. This is seen in Figure 3.5, where we plot the changes in emission energy against the corresponding change in electric field, measured from either the *p*CNF probes or the GFP fluorophore. Figures 3.5A and 3.5B show the changes in emission energy plotted against the field changes projected onto the electronic and vibrational chromophores, respectively, for the mutants containing *p*CNF at position 165. Similarly, Figures 3.5C and 3.5D show the same data measured for the mutants containing *p*CNF at position 145. As can be seen in Figures 3.5A-3.5D changes in the observed fluorescence are strongly correlated with changes in electric field measured from both *p*CNF probes and the fluorophore, with $r >$

0.91 for all comparisons. The size of the error bars is not indicative of variability in the peak centers of the fluorophores, which remained remarkably constant across replicate measurements (standard deviations were ~0.05 cm^{-1} for all *p*CNF probes and ~0.1 nm for the GFP fluorophore, determined by at least three measurements of each). Rather, the size of the error bars is a result of both the uncertainty in the calibrated Stark tuning rates and the resolution of the UV-Vis and FTIR spectrometers. The difference in sign on the slopes of the two plots in Figures 3.5B and 3.5D, measured from the *p*CNF probes, as well as their different magnitudes along the x-axis, highlights the spatial dependence of energy changes based on Equation 2. Because *p*CNF 145 and *p*CNF 165 are oriented differently in space (Figure 3.2), projections of the same field onto their respective bond axes are not expected to give the same value or sign of field, which in this case is revealed by the differing signs of the slopes.

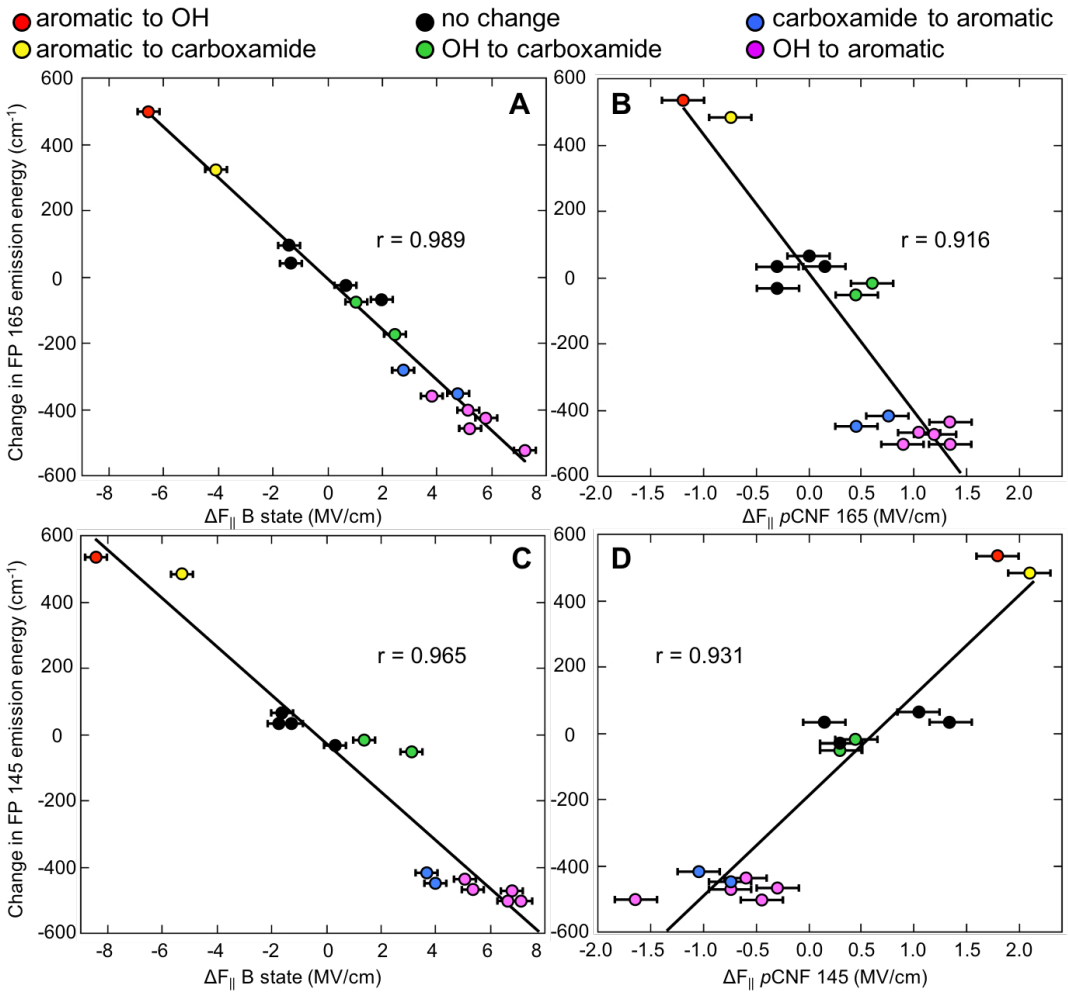

Figure 3.5: Change in the emission energy of the intrinsic fluorophore plotted against the change in field, calculated from B state absorption energies (A, C) or *p*CNF absorption energies (B, D), in response to mutations at position 203. The 15 data points on each plot are derived from the total number of energy differences between the six T203X mutant energies. The values of $\Delta\vec{F}$ on the x-axes are electric field projections calculated from Equation 1.2. On the y-axes, FP stands for fluorescent protein. Colors represent the type of mutation at position 203. Error bars represent the maximum error based on standard deviations in published Stark tuning rates. The slopes of the best-fit lines (in $cm^{-1}/(MV/cm)$) are A: -76.3, B: -355.2, C: -72.8, D: 303.2.

In addition to the strong correlation of color and electric field from each location, the data were grouped into clusters based on the molecular identity of the mutations. Of the six amino acid side chains that we placed at position 203, we grouped each based on being aromatic (Phe, Tyr, and His), polar with an –OH group (Ser and Thr), or polar with a carboxamide group (Asn). The color scheme in Figure 3.5 represents different types of mutations based on this grouping of side chains. It may appear as though the data points are expanded artificially by plotting opposite types of mutations on the same plot (OH to aromatic vs. aromatic to OH, e.g.). However, we should note that in Figure 3.5 (and the following figures) the 'aromatic to OH' and 'OH to aromatic' data points are not simply derived from the same mutation with the order of subtraction reversed. The data points arising from the apparent order change are unique, for example: Tyr to Ser versus Thr to Phe. These two mutations at position 203 represent an apparent change in the order of subtraction. However, it is not clear *a priori* that these two mutations should cause approximately the same magnitude shift in the opposite direction, and thus the observation that they do is significant. Furthermore, it is possible to change the data sets such that the subtraction is done in the same way every time (e.g. only OH to aromatic). However, doing this does not result in any significant changes to the trends that we see (data not shown), nor does it have any effect on the conclusions we draw from these studies.

Cataloging His and Tyr by this scheme was somewhat ambiguous due to the presence of an –OH group on Tyr and the possibility of His to be charged and not aromatic. However, we reasoned that because mutations involving these side chains were closer in energy to aromatic mutations (vida infra), His 203 is not charged and the aromatic ring is more important than the hydroxyl group for the interaction between the fluorophore and Tyr 203. Indeed, previous measurements of the pKa of His 203 have

indicated that it is not charged at neutral pH.[109] Figure 3.5 clearly shows that when grouped in this way, similar types of mutations caused similar changes in field and emission energy regardless of which chromophore (electronic or vibrational) was being used to measure the field change. These results unambiguously demonstrate that the two Stark chromophores responded identically to the same electrostatic perturbations caused by mutations at position 203, independent of the exact transition being measured or position of the probe within the < 10 Å diameter region we investigated (Figure 3.1). This further suggests that the non-native *p*CNF probes can be inserted near the GFP fluorophore without disrupting its intrinsic sensitivity to nearby mutations.

3.5: Electrostatic Perturbation Due to *p*CNF Probes

Interest in using the nitrile as a vibrational Stark probe has occasionally been tempered by the worry that the large ground state dipole moment of the oscillator itself (~4 D) could disrupt the local electrostatic environment of a biomolecule to such a great extent that the label could invalidate the measurement. A necessary step in the interpretation of nitrile frequencies in terms of electric field changes is the quantification of the perturbation due to the nitrile-containing probe, which is often indirectly assessed through functional assays or thermal stability measurements. If an inserted nitrile probe does not greatly alter protein function or stability, measured through a thermodynamic or kinetic parameter, then it is assumed that the environment around the probe must not have been very affected by the insertion of the probe. While the stability and function of a protein are undoubtedly linked to local electrostatic environment, they are not direct reporters of it. Even in the instance that a crystal structure can be determined for both native and probed proteins, electrostatic information is exceedingly difficult to interpret from structural data.[33]

With GFP, the presence of an intrinsic fluorophore formed through protein folding provides an independent handle to investigate any possible alteration of the electrostatic environment caused only by the insertion of the *p*CNF probes themselves. We therefore measured the changes in field experienced by the GFP fluorophore in response to the insertion of *p*CNF at either position 145 or 165. Figure 3.6 shows the magnitude of this field change (calculated from Equation 1.2) for each of the six T203X variants when *p*CNF was inserted at either position 145 or 165. For nearly all of the T203X mutants, the insertion of a nitrile probe at either position induced a change in field of up to ±3 MV/cm relative to the wild type fluorophore. However, when we compared the changes in fluorophore emission energy to changes in field measured by the electronic fluorophore with and without the nitrile probes (Figure 3.7), it is clear that the nitrile probes do not affect the intrinsic fluorophore's sensitivity to the T203X mutations. In this instance, the changes in field experienced by the fluorophore due to the T203X mutations (circles) are independent of either inserted nitrile probe (squares or diamonds). Furthermore, the same patterns of functional groups of the side chain at position 203 were preserved compared to Figure 3.5. These results together suggest that *p*CNF probes, while causing changes in the absolute value of the electric field in their immediate environment, do not affect the response of the native protein to changes in field caused by nearby mutations. This result is significant, because while the determinants of the nitrile oscillation frequency may be too complicated for it to report accurately about individual environments, it appears to be an unobtrusive reporter of changes between two similar environments. The latter of which is by far the most common application of vibrational Stark effect spectroscopy.

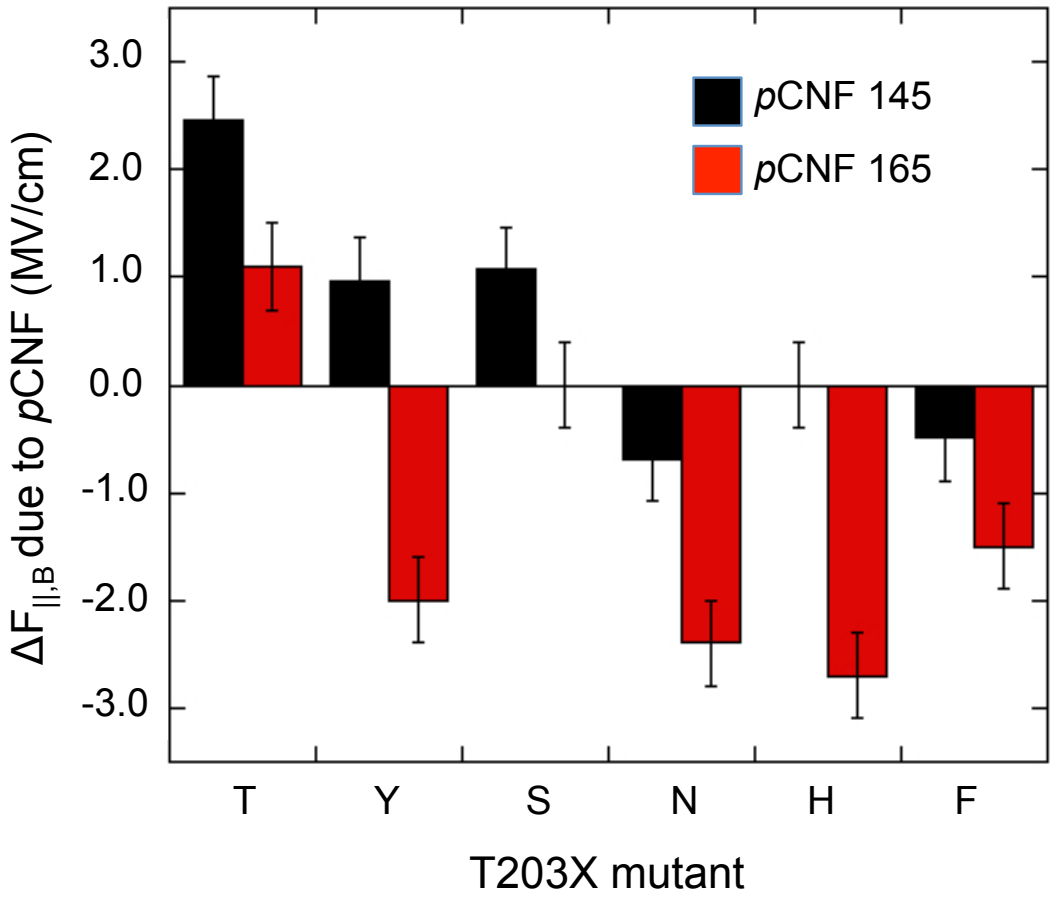

Figure 3.6: The change in field experienced by the B state of the GFP fluorophore, due to the insertion of *p*CNF probes at either position 145 or 165. The x-axis denotes the identity of the residue at position 203 by its one letter amino acid code. Error bars represent the maximum error based on standard deviations in published Stark tuning rates.

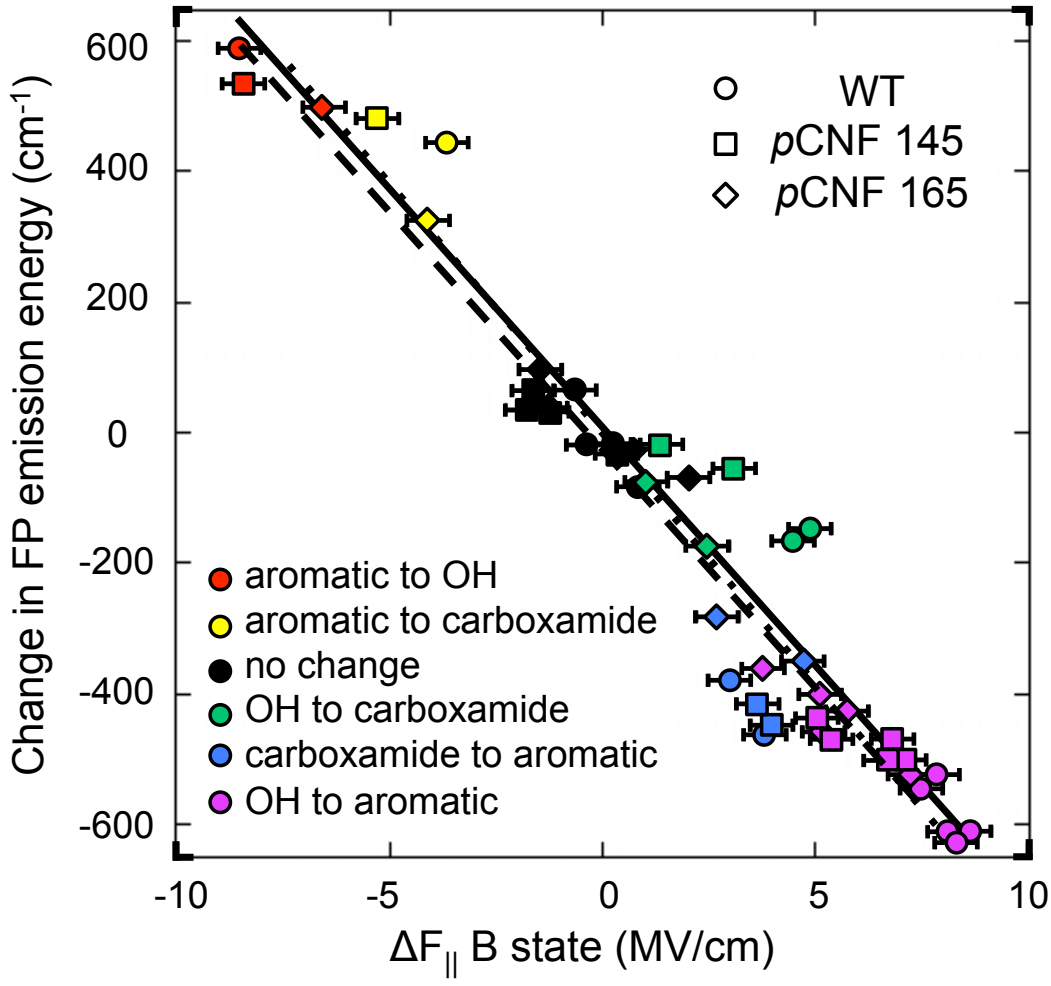

Figure 3.7: Change in emission energy due to position 203 mutation plotted against the change in field of the B state of the fluorophore, with (squares and diamonds) and without (circles) *p*CNF probes. The values of $\Delta\vec{F}$ on the x-axis are electric field projections calculated from Equation 1.2. On the y-axis, FP stands for fluorescent protein. Colors represent the type of mutation at position at 203. Best-fit lines are shown in black for wild type (solid line; $r = 0.960$), *p*CNF 145 (dashed line; $r = 0.965$), and *p*CNF 165 (dotted line; $r = 0.989$). Error bars represent the maximum error based on standard deviations in published Stark tuning rates.

3.6: Vibrational and Electronic Probes Agree Even in the Presence of Hydrogen Bonding

The final issue we sought to investigate with the comparison of vibrational and electronic chromophores is the effect of hydrogen bonding on the vibrational Stark effect spectra of nitriles. There has been significant recent debate over the utility of nitrile vibrations as electric field reporters in the presence of hydrogen bonding, which is known to shift the oscillator vibration to higher energy.[63,68,72,73,110–112] It has been demonstrated that this shift can be deconvoluted into hydrogen bonding and Stark effects in water using independent ^{13}C NMR measurements in concert with vibrational spectroscopy; however, this has not been demonstrated in solvents other than water.[73] Because of this, it has been suggested that because of their predominate sensitivity to hydration, the interpretation of nitrile frequencies in terms of electric fields via Equation 1.2 requires that the nitrile be free of hydrogen bonding.[113] If true, this would significantly limit the utility of nitriles through the vibrational Stark effect in biological systems. GFP is an excellent example of this; although the nitrile chromophore was inserted at positions in the protein interior, multiple crystal structures of GFP clearly reveal significant quantities of water inside the beta-barrel of the protein, some of which is shown in Figure 3.1. If the nitrile chromophore cannot be used in such a nominally sequestered environment, its utility in general will be very limited.

To investigate this problem in the GFP system, in Figure 3.8 we have compared the changes in electric field, measured from both of the *p*CNF probes and the intrinsic fluorophore absorption, in response to the T203X mutations. If hydrogen bonding to the nitrile probes rendered them ineffective as electric field probes, then we would not expect good correlation between the measured frequencies of the vibrational and electronic chromophore absorptions because additional physical effects besides electric field would

be convoluted into energy shifts of the nitriles. However, Figure 3.8 shows that from both nitrile probe locations, we observed a monotonic trend between the changes in field measured from the vibrational reporter and those from the electronic reporter with high correlation constants ($r > 0.87$ in both cases). This finding suggests one of two possibilities: a) that neither of the *p*CNF probes formed hydrogen bonds with water or nearby protein atoms, or b) that one or both of the *p*CNF probes did engage in hydrogen bonds but that this did not affect the ability of either of these probes to report on changes in electric field. While it seems likely, by inspection of crystal structures (Figure 3.1), that a *p*CNF inserted at either location 145 or 165 engages in hydrogen bonding due to the abundance of nearby water molecules at these positions, it is worthwhile to further investigate the hydrogen bonding status of nitriles at these positions due to the possible implication of this for the interpretation of Stark effect spectra.

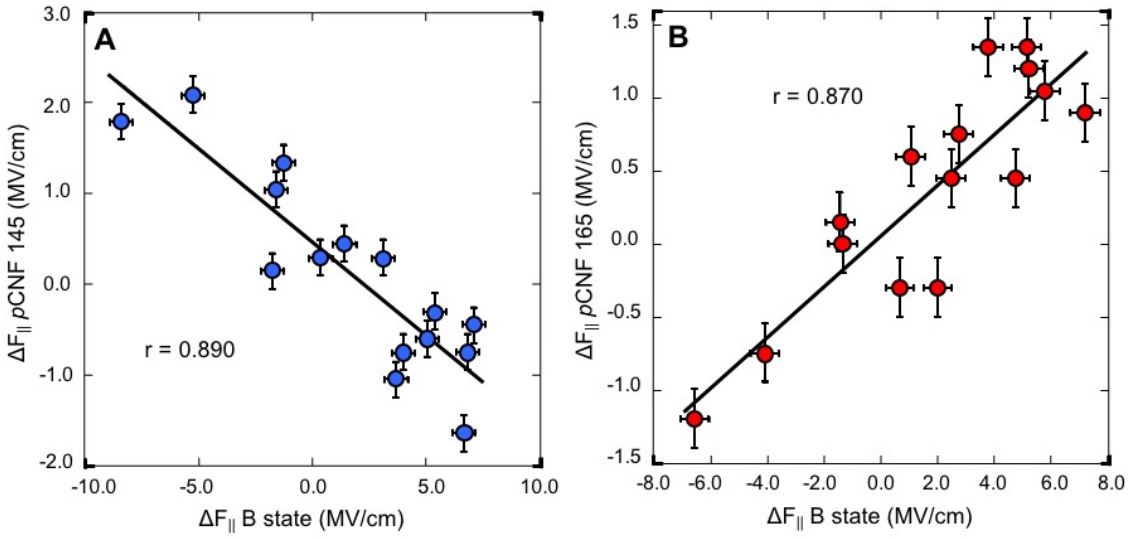

Figure 3.8: Comparison of changes in field measured from *p*CNF probes (A: 145; B: 165) and B state of GFP fluorophore). The values of $\Delta\vec{F}$ on the x- and y-axes are electric field projections calculated from Equation 1.2. Error bars represent the maximum error based on standard deviations in published Stark tuning rates. The slopes of the best-fit lines are A: -0.206 and B: 0.173.

Recently, Adhikary et. al. have reported a straightforward method for determining the hydrogen bonding status of nitrile probes based on the temperature dependence of vibrational frequencies, or frequency temperature line slope (FTLS).[113] It was shown that the magnitude of the FTLS is unique for small nitriles in a variety of protic and aprotic solvents, and that the model nitrile data can be used to assess the hydrogen bonding status of nitrile probes inserted at various locations in a protein. Here we used the same analysis for *p*CNF, *p*-tolunitrile, and GFP mutants containing *p*CNF at positions 145 and 165. Figure 3.9 shows the change in frequency as a function of temperature for small nitriles in three solvents, as well as *p*CNF 145 and 165 for the T203S GFP mutants. Again, the error bars are indicative of the instrument resolution and not the variability across multiple measurements. Similar to previous reports, we observed characteristic FTLSs of *p*-tolunitrile and *p*CNF when dissolved in an aprotic solvent (THF), a moderate hydrogen bonding solvent (formamide), and water, with the magnitude of the FTLS increasing in that order. These data were then compared to the temperature dependent spectra of two constructs of GFP containing a *p*CNF residue at either position 145 or 165. For both nitrile locations, the FTLSs were very similar to that of the small molecule in formamide, which suggests that they experienced a moderate amount of hydrogen bonding in GFP. This observation for the position 145 nitrile is interesting, because its low frequency would ordinarily suggest a lack of hydrogen bonding based on the well-known shift to higher frequency upon the acceptance of a hydrogen bond.[31,32] However, Cho, et. al. have shown that the nitrile frequency can span ~25 cm^{-1} depending on the geometry of the hydrogen bond.[33] It seems likely to us that *p*CNF 145, being shielded from the solvent and exposed to confined water molecules, might be forced to adopt a hydrogen bonding geometry that gives rise to the observed low frequencies. Further high resolution structural studies will hopefully provide some insight into this interesting observation.

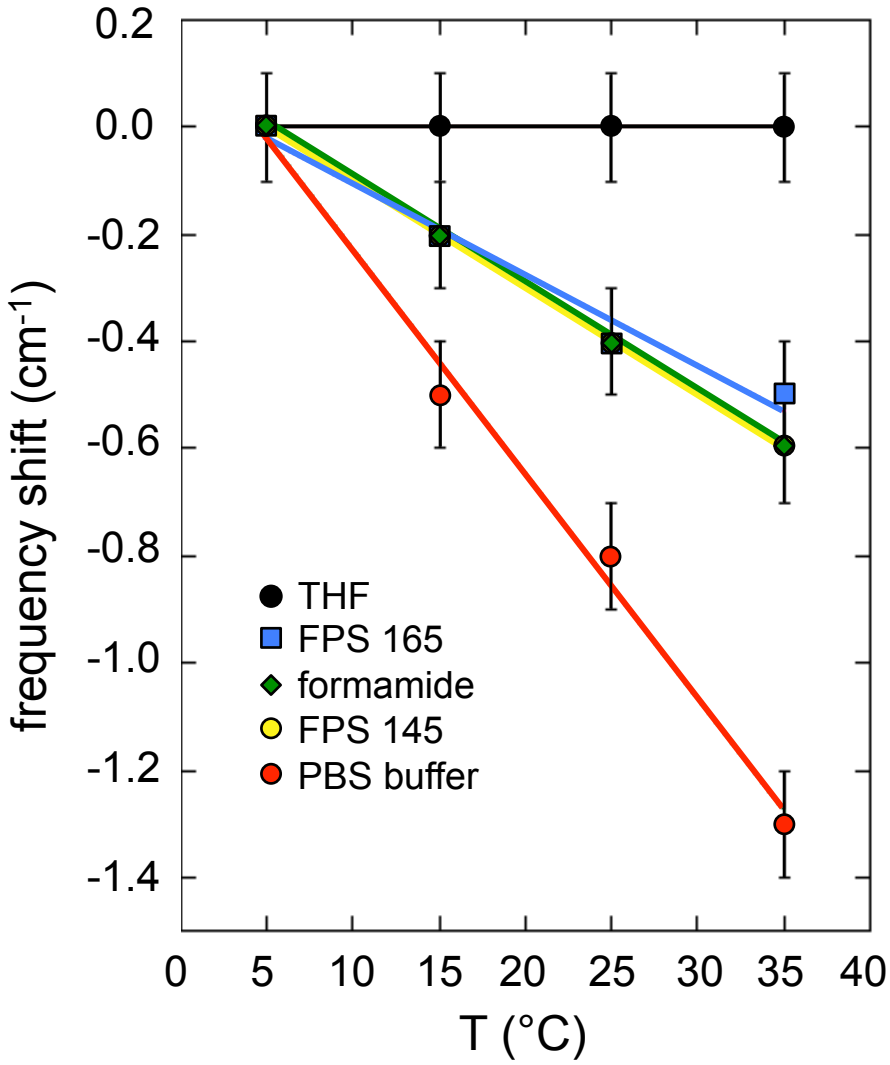

Figure 3.9: Changes in nitrile stretching frequency (relative to the value at 5 °C) as a function of temperature. Black, green, and red data points represent the frequency of *p*-tolunitrile in the noted solvent (or *p*CNF, in the case of PBS buffer). Blue and yellow data points are GFP T203S constructs that contain *p*CNF at position 165 or 145, respectively. Error bars represent the instrument resolution. Standard deviations across at least three measurements were ~0.05 cm^{-1} for all nitriles.

To rule out the possibility of subtle structural changes to GFP causing the observed frequency shifts, we recorded circular dichroism spectra of WT and *p*CNF 165 over this small temperature range (Figure 3.10). Over the range measured GFP did not have any measurable change in secondary structure, which suggests that the observed FTLSs of *p*CNF 145 and 165 are indicative of moderate hydrogen bonding. This result supports our hypothesis from crystallographic data that nitriles at these two positions are likely to engage in hydrogen bonding, probably from water molecules contained within the beta-barrel structure of GFP. In the crystal structure shown in Figure 3.1, as well as others that have been solved for similar constructs of GFP, positions 145 and 165 are entirely shielded from bulk solvent but are in close proximity to several confined water molecules. Together these data suggest that *p*CNF probes at these locations participate in hydrogen bonding, and yet still accurately report on changes in field as measured from the electronic fluorophore.

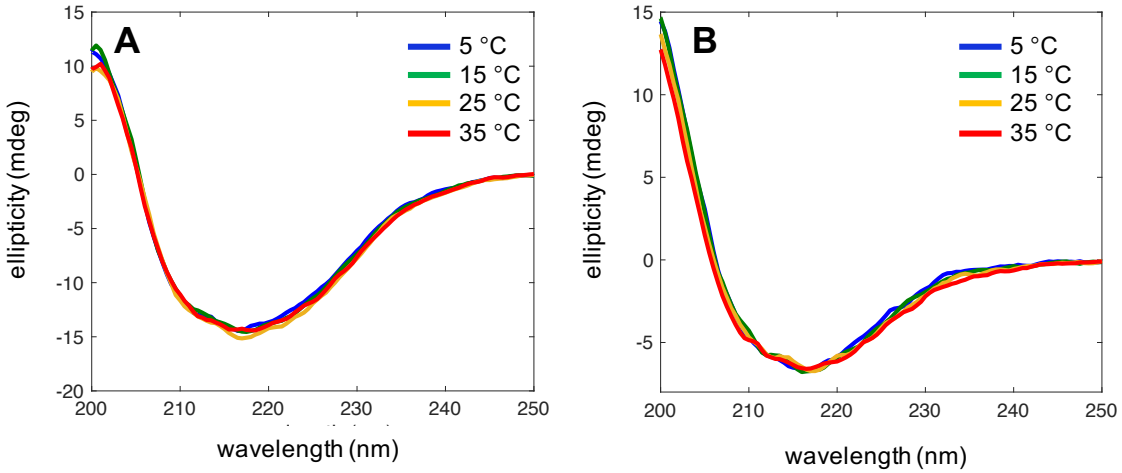

Figure 3.10: Circular dichroism spectra of mutants containing *p*CNF 145 (A) and *p*CNF 165 (B) between 5-35 °C.

This finding is significant because it suggests that it is possible to relate changes in the absorption energies of nitrile probes to changes in electric field, even in the presence of hydrogen bonding. This result is similar to the conclusion reached by Fafarman, et. al., where nitrile probes exposed to bulk water had frequencies that were all shifted by a constant 10 cm^{-1} with respect to an independent NMR parameter.[73] In the present study, there also appears to be a constant contribution to the nitrile frequency from the hydrogen bonds to ordered water molecules that cancels out when comparing two similarly hydrogen bound nitriles. This suggests that small changes in the hydrogen bonding status of the nitrile probe between two states could still cause significant deviations from the Stark effect. In order to observe this independence of the frequency change on hydrogen bonding, it may still be necessary to compare states that are highly similar with respect to the hydrogen bonding status of the nitrile. Further studies are underway to compare Stark effects of the GFP fluorophore to nitrile frequencies of *p*CNF probes inserted into different positions with potentially different hydrogen bonding environments to assess the generality of this finding.

3.7: Conclusion

We have used GFP to explore the behavior of the nitrile vibrational Stark probe in comparison to the known visible Stark behavior of GFP's intrinsic fluorophore. Nitrile probes were inserted at buried locations near the fluorophore without affecting the sensitivity of the observed fluorescence to electrostatic changes made in the vicinity of the fluorophore by amino acid mutations. Additionally, changes in both the intrinsic fluorescence and site-specific electric fields were related to chemical properties of the amino acids involved in the mutation. We have also shown that the changes in absorption frequency of the two *p*CNF probes inserted at positions 145 and 165 caused by these

same amino acid mutations resulted in changes in electric field that correlated highly with electric fields measured independently from the electronic fluorophore.

A series of measurements of the temperature dependence of the nitrile frequencies suggests that they participate in hydrogen bonds with ordered water molecules inside the beta-barrel. This result is not surprising, but is interesting because hydrogen bonding has been thought to complicate the interpretation of nitrile frequencies in terms of electric fields. However, pCNF probes at these two locations appear to accept hydrogen bonds, while simultaneously having absorption frequencies that are well described by the vibrational Stark effect. This suggests that it may not be necessary that a nitrile be free of hydrogen bonds in order to be useful as a Stark effect probe as long as differences in absorption energy between similar states are being considered. While GFP is almost unique in that the protein carries an intrinsic, easily measured electrostatic probe, the conclusions of this study can be generalized to nitriles used as vibrational probes in other proteins.

Chapter 4: Comparing Nitrile Probes of Electric Field to GFP Fluorophore Equilirbrium

4.1: Publication Note

Portions of this chapter were adapted from:

J. D. Slocum, J. T. First, L. J. Webb, "Orthogonal Electric Field Measurements near the Green Fluorescent Protein Fluorophore through Stark Effect Spectroscopy and pK_a Shifts Provide a Unique Benchmark for Electrostatics Models." *J. Phys. Chem. B, 121 (28)*, 6799-6812 (2017). [J. D. Slocum performed all experimental measurements, including protein expression and purification, infrared and UV-Vis spectroscopic measurements, and pK_a measurements].

4.2: Introduction

Measurement of the magnitude, direction, and functional importance of electric fields in biomolecules has been a longstanding experimental challenge. pK_a shifts of titratable residues have been the most widely implemented measurements of local electrostatic environment around the labile proton, and experimental data sets of pK_a shifts in a variety of systems have been used to test and refine computational prediction capabilities of protein electrostatic fields. A more direct and increasingly popular technique to measure electric forces in proteins is Stark effect spectroscopy, where the change in absorption energy of a chromophore relative to a reference state is related to the change in electric field felt by the chromophore. While there are merits to both of these methods and they are both reporters of local electrostatic environment, they are fundamentally different measurements, and to our knowledge there has been no direct comparison of these two approaches in a single protein. Measurements of the direction and magnitude of electric field changes in proteins could provide a convenient complement to pK_a measurements of titratable residues for researchers who wish to

benchmark electrostatics models. In principle, because pK_a values are determined in part by noncovalent interactions, and site-specific electric fields represent the force at a particular location due to noncovalent interactions, these two types of measurements should give the same result when used in the same system. However, they represent an orthogonal set of measurements which can be used to describe electrostatic effects in fundamentally different ways, and to our knowledge, there has been no direct comparison of pK_a and Stark effect measurements in the same system. It is the goal of this work to compare pK_a shifts to measured electric field changes, caused by the same perturbations in a protein.

Recently, we reported measurements of electric fields in super folder green fluorescent protein (GFP), inferred from Stark shifts of biosynthetically incorporated nitrile probes and from the GFP fluorophore, in response to a series of amino acid mutations.[114] We observed an agreement between these two types of Stark effect probes, which provides strong confidence for interpreting the spectroscopic responses of these probes in terms of electric fields. Because the GFP fluorophore, shown in Figure 4.1, can be protonated or deprotonated at physiological pH, it provides a convenient pK_a probe in the vicinity of the Stark effect probes.

Here we report the pK_a values of this fluorophore, estimated from the pH-dependent absorption spectra of the GFP fluorophore, as a function of the same mutations. To do this, we inserted the unnatural amino acid, *p*-cyanophenylalanine (*p*CNF), at positions 145 and 165, which are both near the GFP fluorophore, and made a series of mutations to position 203 (Figure 4.1). We then measured the effect of these mutations on the electric field in the vicinity of the flourophore both through VSE spectroscopy of the two nitrile probes and pK_a measurements of the flourophore itself. Our results show that the pK_a shifts in response to the amino acid mutations are linearly

correlated with the site specific electric field changes, measured from either the vibrational or electronic Stark probes. This agreement provides confidence in both of these types of experimental measurements and also allows the pK_a shifts to be broken down into site-specific electric field changes. This trend was not observed when histidine or cysteine were at position 203, which we hypothesize was due to (1) differences in the hydrogen bonding network around the fluorophore caused by histidine and cysteine at position 203, or (2) the interference of the fluorophore titration by the simultaneous titration of histidine or cysteine at position 203. All together, this model system contains a vibrational Stark effect probe, an electronic Stark effect probe, and a pK_a probe that all respond similarly to perturbations at position 203. In addition to providing confidence in our ability to measure electrostatic effects experimentally, we also believe that these experimental data are an excellent target data set for benchmarking a variety of computational electrostatics models. A significant portion of the full manuscript includes work by Jeremy First, using molecular dynamics simulations and continuum electrostatics models to recreate the measured values reported here. Those data will not be presented here.

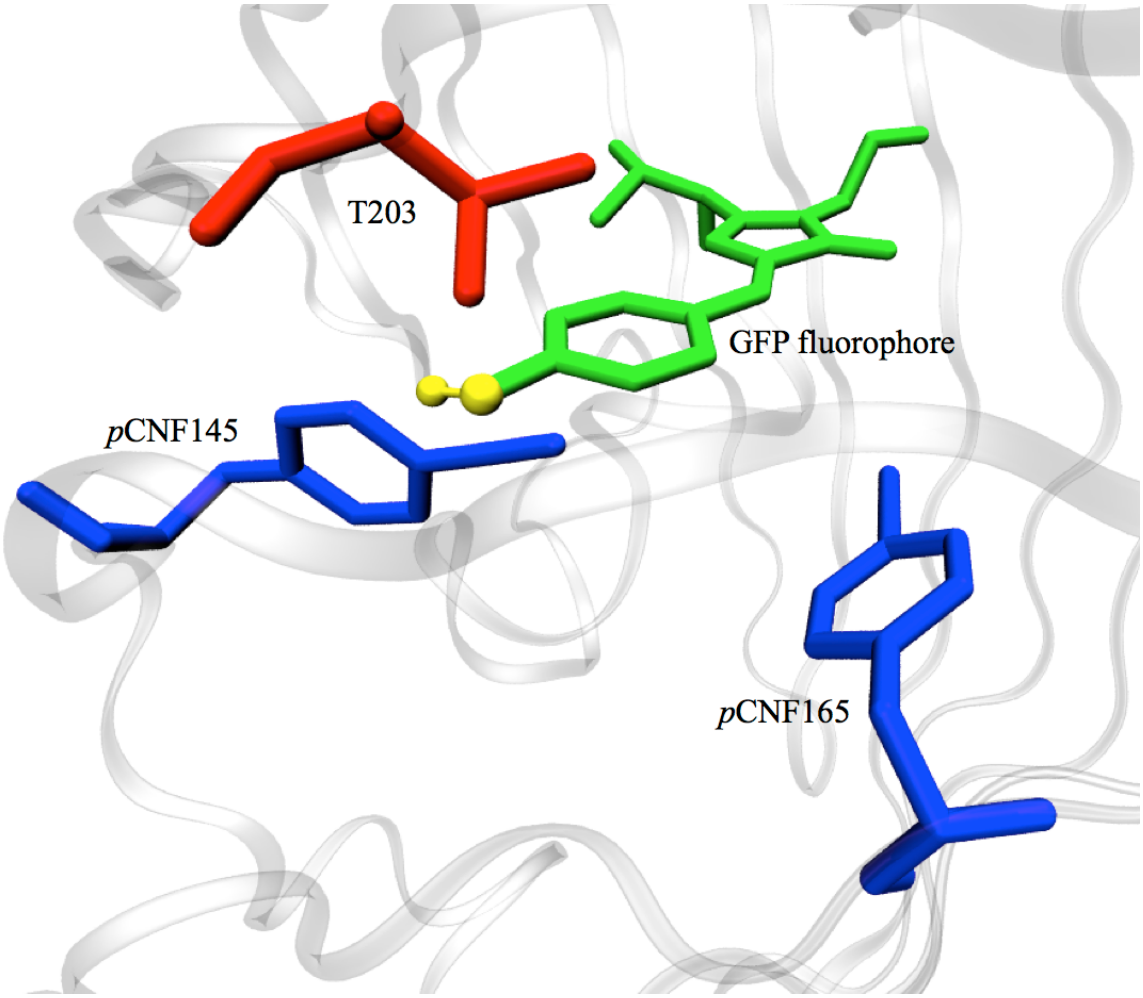

Figure 4.1: A close-up view of the interior of GFP (gray ribbon) modelled from crystal structure 2b3p. The GFP fluorophore is shown in green with the titratable –OH shown in yellow. *p*CNF 145 and 165 are shown in blue and position 203 is shown in red.

4.3: pK_a Measurements

To compare the ability of pK_a shifts and Stark shifts to report on changes in electrostatic environment in a protein, we mutated the wild type T203 (where T represents the one letter code of threonine and 203 is the amino acid position in the primary sequence) of GFP to S, N, H, C, F, and Y and inserted a *p*CNF residue at either position 145 or 165 (Figure 4.1). To estimate the equilibrium between the neutral and ionized fluorophore, we measured the pH dependence of the visible absorption of the fluorophore over a range of pH 5-10. Figure 4.2 shows the pH-dependent visible absorption spectra of a representative GFP mutant, S203 (Figure 4.2A), and the resulting titration curve (Figure 2B), from which the pK_a values were determined. As expected, the fluorophore population shifted from mostly protonated (A state) at low pH to mostly de-protonated (B state) at high pH. The presence of a clean isosbestic point indicated that there were only two states participating in the equilibrium. Plotting the maximum absorbance of both the A and the B states as a function of the solution pH resulted in two curves which could be fit to a sigmoidal function and used to find the inflection point. The curve fitting application of Matlab R2016a was used to fit all titration curves to a sigmoidal function of the form shown in Equation 4.1,

$$Abs(pH) = \frac{a}{b + e^{c \cdot pH}} + d \qquad \text{[Equation 4.1]}$$

where *Abs*(*pH*) is the maximum absorbance of either the A or B state at a given *pH*, *a* is the maximum value of *Abs*, *b* defines the offset of the function from *pH* = 0, *c* defines the pitch of the sigmoid, and *d* is the minimum value of *Abs*.[115] The inflection point of the resulting fit was taken to be the pK_a for that mutant. The titration curves for each of the 21 mutant constructs is shown in Figure 4.3. With the exception of the few cases where the A state titration could not be fit (Y203, H203, and H203 with *p*CNF 165) in Figure 4.3, titrations of both the A and B states gave pK_a values that were within

experimental error of each other, < 0.08 pK_a units. In the cases where only one titration could be fit to a sigmoidal curve, the pK_a was estimated directly from that titration.

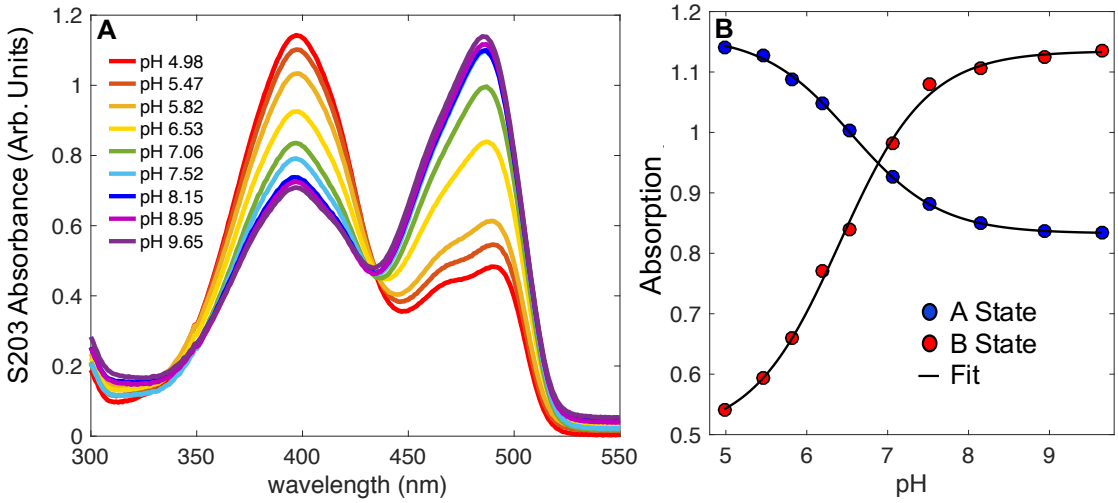

Figure 4.2: A representative titration of the GFP fluorophore. Panel A shows the visible absorption spectra of the fluorophore of S203 over a range of pH values given by the color code. Panel B shows the maximum absorption of the A state (blue) and the B state (red) of the S203 spectra from Panel A, plotted against the solution pH. The solid lines are sigmoidal fits which are used to estimate the pK_a.

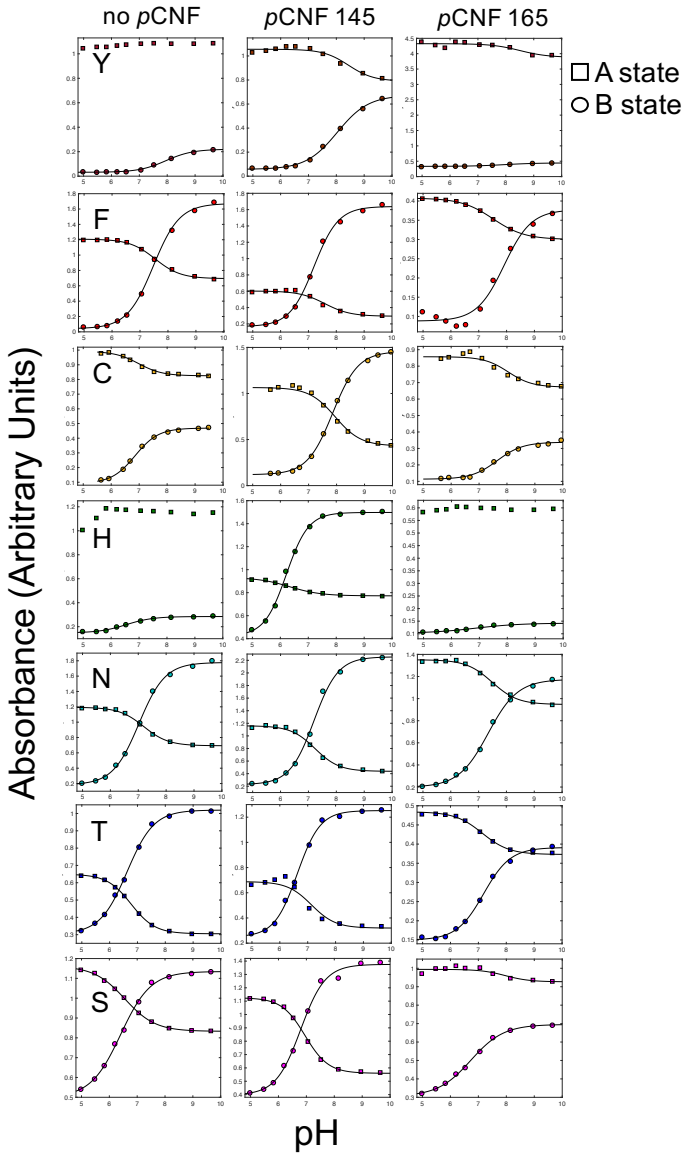

Figure 4.3: Titration curves for all 21 mutants. The one letter codes in the first column designates the identity of the amino acid at position 203 for each row. The three columns are for mutants that contain no *p*CNF, *p*CNF 145, or *p*CNF 165 as indicated above each column. Squares represent the A state titrations and circles represent the B state titrations.

Table 4.1 shows the pK_a values obtained for all 21 of the mutant constructs. Overall, we measured a range of pK_a values from 6.28 – 8.04, which is typical for GFP mutants containing the TYG fluorophore. Previous studies have suggested that two dominating contributions to GFP fluorophore pK_a values are (1) specific hydrogen bonding to or from the fluorophore phenolate (or phenol), determined by the local hydrogen bonding network of water and protein residues around the fluorophore, and (2) aromatic interactions with the fluorophore from nearby side chains that tend to stabilize the neutral form of the fluorophore.[91,92] Here, the measured pK_a values tended to increase as we placed aromatic residues at position 203, and the lowest pK_a values were measured for the small, polar residues at position 203. This suggests that aromatic interactions between the fluorophore and Y and F at position 203 are responsible for the elevated pK_a values we have measured here, and that the hydrogen bonding environment around the fluorophore might play a larger role in the pK_a values for the other position 203 mutants. To further investigate these interactions and their importance in determining fluorophore pK_a, we compared measured pK_a changes to the orthogonal site-specific electric field measurements obtained through spectroscopic Stark effect measurements using both the GFP fluorophore and the biosynthetically incorporated *p*CNF residues.

Pos. 203	WT	*p*CNF 145	*p*CNF 165
T*	6.70 ± 0.07	6.63 ± 0.04	7.17 ± 0.03
F	7.50 ± 0.03	7.20 ± 0.04	7.51 ± 0.05
C	6.85 ± 0.05	7.86 ± 0.04	7.65 ± 0.04
H	6.54 ± 0.03	6.28 ± 0.05	6.77 ± 0.05
N	7.11 ± 0.08	7.22 ± 0.04	7.42 ± 0.06
S	6.48 ± 0.07	6.88 ± 0.07	6.73 ± 0.05
Y	7.95 ± 0.04	7.95 ± 0.04	7.95 ± 0.03

Table 4.1: A summary of the measured pK_a values for all GFP mutants. Columns are the different nitrile locations: WT (no nitrile), *p*CNF at position 145, and *p*CNF at position 165. The rows designate the different amino acid side chains at position 203 based on the one letter codes in the first column. *The identity of position 203 in the wild type protein is T.

4.4: Vibrational and Electronic Absorption Measurements

To measure the local electric field changes near the fluorophore due to the position 203 mutations, we measured the absorption spectra of the GFP fluorophore and the inserted *p*CNF probes. Figure 4.4 shows representative spectra of all variants investigated. Figures 4.4A, 4.4B, and 4.4C show the fluorophore absorption as a function of the position 203 mutations when there is no *p*CNF, *p*CNF at position 145, and *p*CNF at position 165, respectively. All visible spectra showed the same general shape, with two major features which have been well characterized as being due to the neutral (A state: ~400 nm) and ionized (B state: ~500 nm) forms of the fluorophore. There were two distinct types of changes in these spectra that resulted from the position 203 mutations. First, changes in the relative intensities of the two major peaks indicated that the equilibrium between neutral and ionized fluorophore changed (for example, Y and T in Figure 4.4A, which show different ratios of absorbance between the A and B states). Second, the energies of maximum absorption of both the A and B state changed in response to the position 203 mutations. The B state absorption energies, which have been shown to be well described by Equation 1.2, with $\Delta\vec{\mu}$ = 117.5 $cm^{-1}/(MV/cm)$[39] for absorption of the electronic fluorophore, will be used in the following sections to calculate electric field changes experienced by the fluorophore. However, the A state absorption energies, which still changed as a function of the mutations, are not well described by the linear Stark effect and cannot be straightforwardly related to electric field changes.[39] Figure 4.4D shows the representative *p*CNF absorption spectra of *p*CNF 145 (solid) and 165 (dashed). With the exception of the mutant containing C203 and *p*CNF 145 (Figure 3D, solid yellow line), all of these spectra could be well described by a single Gaussian whose center frequencies spanned a range of approximately 2 cm^{-1} across all mutants of position 203. As we previously reported, the *p*CNF 145 spectra were quite

narrow (~5-6 cm^{-1} FWHM, except C203) and the *p*CNF 165 spectra were somewhat more broadened (~8-9 cm^{-1}), suggesting that the nitriles at position 165 experience a more heterogeneous environment than the nitriles at position 145. In addition to this, the *p*CNF 145 and 165 spectra were separated from each other in energy by ~10 cm^{-1} depending on the identity of the side chain of position 203. Altogether, these spectra suggest that the local environment around position 145 is quite different than that of position 165, despite their proximity to one another (Figure 4.1). This is likely a reflection of the environmental complexity of the interior of GFP, which could be manifested in different hydrogen bonding environments or local conformational fluctuations of the nitriles, and will be discussed in more detail below.

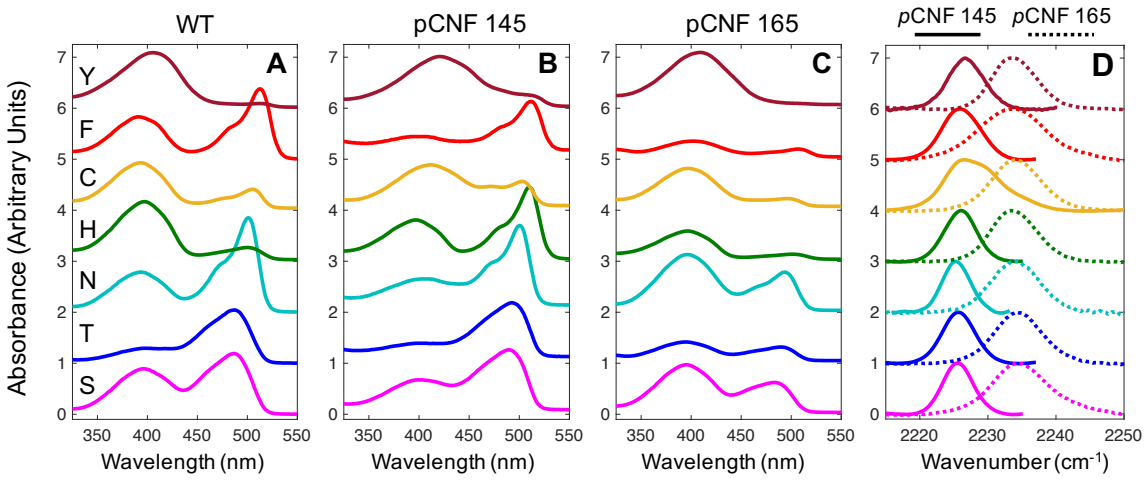

Figure 4.4: Representative absorption spectra for all GFP mutants. A, B, and C show the visible absorption of the GFP fluorophore normalized to the absorbance at 280 nm for proteins that have no nitrile probe (A), *p*CNF at position 145 (B), and *p*CNF at position 165 (C). Different colors represent the different amino acid side chains at position 203 based on the one letter code in panel A. Panel D shows the FTIR absorption of the *p*CNF residue at either position 145 (solid) or 165 (dashed) with the same colors to denote different amino acids at position 203. All spectra have been offset on the y-axis for clarity.

4.5: Comparing pK_a Shifts to Electronic Stark Effect Shifts

To compare the changes in fluorophore pK_a to changes in electric field experienced by the fluorophore, we plotted the pK_a changes for all unique pairs of the seven mutants against the corresponding changes in electric field experienced by the fluorophore (calculated from Equation 1.2) in Figure 4.5. The data in Figures 4.5A-C were generated from GFP constructs that contained no nitrile probe (Figure 4.5A), *p*CNF at position 145 (Figure 4.5B), or *p*CNF at position 165 (Figure 4.5C). In this case, the seven individual pK_a values and deprotonated fluorophore absorption energies from the position 203 mutants gave rise to 21 distinct pK_a shifts and absorption energy changes. We used these absorption energy changes and the known Stark tuning rate of the deprotonated GFP fluorophore to calculate the electric field changes experienced by the fluorophore from Equation 1.2. For all of these comparisons, we observed a strong linear correlation between the fluorophore pK_a shifts and corresponding Stark effect shifts due to the position 203 mutations, which suggests that these two orthogonal measurements responded similarly to the same electrostatic perturbation. Furthermore, responses to electrostatic perturbations followed previously observed patterns based on the chemical identity of the mutant residue. In Figure 4.5, the different colors represent the type of mutation that was made at position 203 based on the key in Figure 4.5A. For instance, the magenta points represent mutations involving the replacement of an –OH side chain (T or S) with an aromatic side chain (F or Y). We had previously observed that mutations of this nature, and their opposites (red in Figure 4.5), are responsible for the largest changes in electric field, measured spectroscopically. In this work, we see that these mutations also cause the largest shifts in pK_a. Additionally, we saw that conservative mutations (e.g. S to T or F to Y; black in Figure 4.5) tended to cause the smallest changes in both pK_a and electric field experienced by the fluorophore. This observation, that the Stark shifts

and pK_a shifts of the fluorophore are not only linearly correlated, but also respond similarly to the same types of mutations (denoted by the colors in Figure 4.5), gives us confidence that these two orthogonal probes of electrostatic perturbation are indeed reporting on a common electrostatic effect.

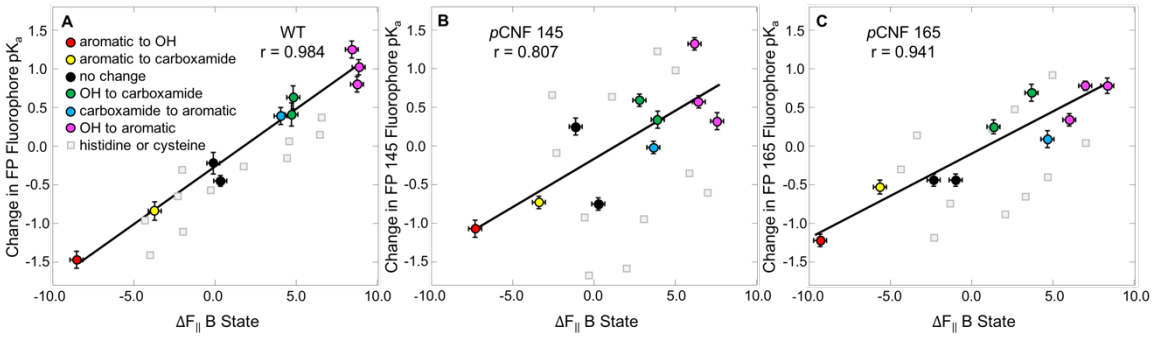

Figure 4.5: Changes in the measured fluorophore pK_a plotted against changes in the electric field experienced by the B state of the fluorophore for GFP mutants with (A) no nitrile probe present, (B) *p*CNF at position 145, and (C) *p*CNF at position 165. Different colors represent the type of mutation that was made to position 203 based on the key in panel A. Gray squares represent position 203 mutations that involved changes to or from histidine or cysteine and are omitted from the best fit lines and correlation constants. Error bars on the x-axis represent uncertainty in the published Stark tuning rate of the deprotonated GFP fluorophore. Error bars on the y-axis represent standard deviations in the inflection points of at least three pK_a titration curves.

In Figure 4.5, however, there are clear deviations from this linear correlation. In displaying the data in this manner, it is apparent that, with the exception of the proteins with no *p*CNF probe (4.5A), mutations to or from histidine or cysteine (gray squares in Figure 4.5) caused pK_a changes that were in poor agreement with the corresponding Stark shifts of the GFP fluorophore. Of the seven side chains we placed at position 203, H and C are clearly different from the rest because they can be titrated near physiological pH (the pK_a values of H and C side chains in water are 6.0 and 8.1, respectively). It is reasonable to suspect that the presence of a second titration site (H or C at position 203) might complicate the straightforward analysis of the GFP fluorophore equilibrium and contribute to the weakening of the correlations shown in Figures 4.5B and 4.5C. However, the fact that the disagreement between electronic Stark effects and pK_a shifts for these mutations is only apparent when either nitrile probe is present suggests that the presence of *p*CNF is the cause of the disagreement. Because of the likely role of hydrogen bonding to the fluorophore in determining its pK_a, and the ability of the nitrile group to accept a hydrogen bond, we think it is likely that *p*CNF at either of these locations could potentially disrupt the hydrogen bonding network in the interior of GFP, which could in turn perturb the native pK_a values. Further studies are underway to investigate changes in the hydrogen bonding network near the GFP fluorophore induced by the insertion of *p*CNF residues at positions 145 and 165 to test this hypothesis. In the sections that follow, we address the electrostatic perturbations induced by the nitrile probes and how they affect our ability to interpret the measured vibrational energies in a meaningful way.

4.6: pK_a Perturbations Due to Nitrile Probes

One of the principal criticisms of using nitrile vibrational probes in biological molecules is that the inclusion of the nitrile itself, an oscillator with a large ground state dipole moment and hydrogen bond accepting ability, will inherently perturb the biomolecule, and is not a reliably benign reporter of the molecular environment of interest.[72] The observation discussed above, that mutations to C or H deviate from the otherwise strong correlation between pK_a and Stark shift when a nitrile probe is present, may indeed be evidence of this effect. To investigate this, we examined the ability of the nitrile probes to perturb the fluorophore pK_a values, relative to the values with no nitrile probe present. This is shown in Figure 4.6, where the change in fluorophore pK_a of all position 203 mutants (indicated by the one letter code on the x-axis) is plotted as a function of a nitrile located at either position 145 (black) or 165 (red). For most of the mutants, we saw that the insertion of a *p*CNF residue at either position induced a pK_a change of at least ± 0.2 pH units, which represents a significant electrostatic perturbation due to the nitrile probe. Because the nitrile moiety can accept a hydrogen bond from either water or a nearby amino acid residue, inserting *p*CNF into the beta-barrel of GFP has the potential to disrupt the hydrogen bonding network around the fluorophore, which could explain the significant pK_a changes in Figure 4.6. C203 mutants with either nitrile probe clearly exhibit the largest observed pK_a shifts, causing shifts of almost 1 pH unit, which further supports our decision to remove C203 from the trends in Figure 4.5. This highlights the care that should be taken when using nitriles as electrostatic probes.

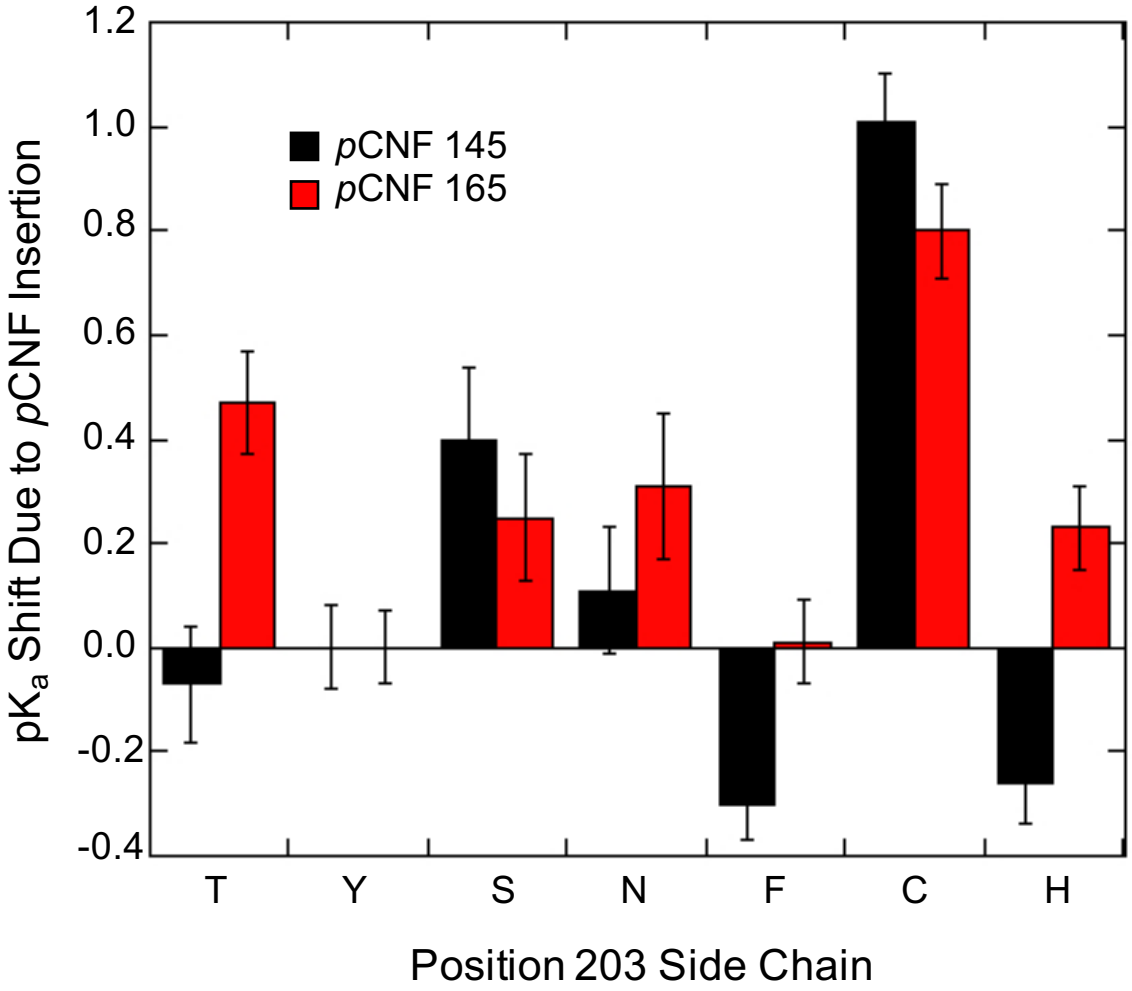

Figure 4.6: The pK_a shift of the GFP fluorophore caused by the insertion of *p*CNF at position 145 (black) or 165 (red) for each amino acid at position 203 (one letter code on the x-axis). Error bars represent the standard deviation of at least three titrations.

In contrast, the *p*CNF residues appear to be least perturbative when Y or F is at position 203. Previous work has suggested that aromatic van der Waals interactions between the residue at position 203 and the fluorophore, as well as specific hydrogen bonding to and from the fluorophore, are the dominating contributions to GFP fluorophore pK_a.[91] As such, we think that the mutants containing Y or F at position 203 have pK_a values that are dominated by aromatic interactions, and thus should be less sensitive to subtle changes in the surrounding hydrogen bonding network. Therefore, with the view that the perturbations shown in Figure 4.6 were caused by differences in the hydrogen bonding network around the fluorophore due to the *p*CNF residues, it makes sense that the smallest perturbations were observed for mutants with Y or F at position 203. This is further supported by the elevated pK_a values we measured for the Y203 and F203 mutants (~7.5-8.0; Table 4.1), which suggest that aromatic interactions between Y203 or F203 and the fluorophore work to stabilize the protonated form of the fluorophore. Interestingly, with the exception of F203 and H203 mutants containing *p*CNF 145, the insertion of the nitrile probe either caused no measurable pK_a change or caused a shift to higher pK_a values (in favor of the neutral fluorophore). While it is difficult to ascribe this general increase in fluorophore pK_a to any specific interactions, it may be the result of favorable aromatic interactions between the fluorophore and *p*CNF 145 or 165, which are in close enough proximity to one another for this to be feasible. Additionally, the inserted nitrile probes might disrupt the hydrogen bonding environment around the fluorophore in a systematic way that tends to favor the neutral fluorophore.

4.7: Comparing pK_a Shifts to Vibrational Stark Effect Shifts

As we have observed before, *p*CNF at these two positions can perturb the native environment around the GFP fluorophore. However, regardless of the mechanism by

which these nitrile probes shift the pK_a of the GFP fluorophore, our primary interest is in relating changes in their measured vibrational energies to changes in electric field via Equation 1.2. We previously observed that, even though *p*CNF induced large electric field perturbations into GFP, changes in the nitrile absorption energy could still be related to electric field changes that were verified by an independent electric field reporter.[114] We hypothesized that a similar effect might happen here, where the observed pK_a perturbations due to *p*CNF 145 and 165 might be cancelled out such that differences in pK_a can still be correlated to VSE shifts of the nitriles. To investigate this possibility, we directly compared the nitrile Stark shifts to the pK_a shifts of the fluorophore, caused by the mutations at position 203. Figure 4.7 shows the fluorophore pK_a changes due to all unique pairs of the seven position 203 mutants plotted against the corresponding electric field changes projected onto either *p*CNF 145 (Figure 4.7A) or 165 (Figure 4.7B) calculated from Equation 1.2 using the nitrile absorption energies as determined from Figure 4.4D. Similar to the data shown in Figure 4.5, we observed a linear correlation between the pK_a shifts and the corresponding electric field changes experienced by the *p*CNF residues, which further strengthens our confidence in the ability of these two orthogonal measurements to report on the same electrostatic perturbations. Additionally, based on the color key in Figure 4.7B, we saw that the same trends in position 203 side chain that were observed from the fluorophore Stark effect response (Figure 4.5) were observed in the response of these nitrile probes. The trend lines in Figure 4.7A and 4.7B have slopes that differ in both sign and magnitude, which highlights the geometric dependence of the vibrational Stark effects shifts (dot product in Equation 1.2). Because *p*CNF 145 and 165 are spatially removed from each other and oriented in different directions (Figure 4.1), we should not expect the projection of an electric field change on to their respective bond axes to result in the same changes in energy.

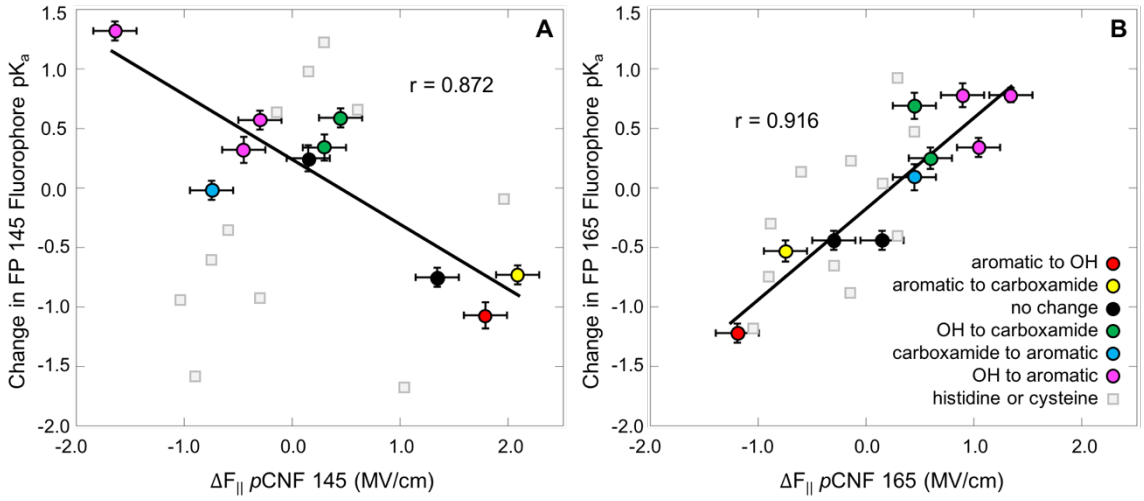

Figure 4.7: Changes in the fluorophore pK_a plotted against the electric field changes measured from *p*CNF 145 (A) and 165 (B) for all unique pairs of the seven position 203 mutants. Colors represent the type of mutation made to position 203 based on the key in panel B. Gray squares represent position 203 mutations that involved changes to or from histidine or cysteine and are omitted from the best fit lines and correlation constants. Error bars on the x-axis represent uncertainty in the published Stark tuning rate of *p*CNF. Error bars on the y-axis represent standard deviations in the inflection points of at least three titration curves.

The observed linear correlation between VSE shifts of these inserted nitriles and the corresponding fluorophore pK_a shifts is remarkable. In the previous section, we showed that *p*CNF 145 and 165 can significantly perturb the fluorophore pK_a values (Figure 4.6), and yet the changes in pK_a are still well correlated to the measured VSE shifts (Figure 4.7). This is similar to what we have previously reported, where *p*CNF residues at these positions in GFP cause a significant electrostatic perturbation to the fluorophore, but do not affect GFP's intrinsic fluorescence sensitivity to the position 203 mutations. The perturbations caused by *p*CNF residues in GFP appear to be systematic, such that the comparison between two states allows any perturbations caused by the experiment itself to be cancelled and the resulting energy difference to be interpreted in a meaningful way. However, again it is clear that mutations to or from cysteine and histidine caused a weakening in the correlation between the pK_a shifts and the electric fields measured from each *p*CNF residue, although the effect is worse for *p*CNF 145 than it is for *p*CNF 165. In summary, we observed that similar types of mutations at position 203 cause similar changes in (1) the vibrational absorption energies of inserted nitrile probes; (2) the visible absorption energies of the GFP fluorophore; and (3) the fluorophore pK_a. The self-consistency in these three orthogonal experimental measurements of electrostatic perturbation from the position 203 mutations makes this experimental dataset an ideal benchmark for computational models that predict electrostatic fields in proteins.

4.8: Conclusion

Here, we have shown that shifts in GFP fluorophore pK_a, caused by nearby amino acid mutations, are in generally good agreement with the corresponding Stark shifts of both the fluorophore and *p*CNF probes inserted near the fluorophore. This observed

agreement between the vibrational and electronic Stark effect probes and pK_a shifts not only instills confidence in each of these measurements as probes of non-covalent environment, but also provides an unprecedented level of information about the electrostatic environment around the GFP fluorophore. Because of this, we think the GFP fluorophore offers a unique benchmark for electrostatics models. Contrary to datasets of pK_a values measured from titratable residues at various sites in the interior of a protein, we have measured pK_a shifts of a single interior site as a function of changes in local environment, caused by the position 203 mutations. In addition to providing a range of pK_a values of the same residue as a function of nearby mutations, this dataset also offers the ability to benchmark electrostatics models against site-specific measurements of electric field changes that arise from the same mutations. In summary, we have reported that electrostatic perturbations, induced by mutations at position 203, cause similar changes in (1) the VSE shifts of inserted nitrile probes; (2) the electronic Stark effect shifts of the GFP fluorophore; and (3) the fluorophore pK_a values. We hope that the data presented here will be useful to other researchers, not only to validate protein electrostatic models, but also to elucidate the underlying physical determinant of fluorophore equilibrium in fluorescent proteins.

Chapter 5: High Contrast Photoactivation in a Point Mutant of Superfolder GFP Caused by Double Decarboxylation of Residues Near the Fluorophore

5.1: Publication Note

Portions of this chapter were adapted from:

J. D. Slocum and L. J. Webb, "A Double Decarboxylation in Superfolder Green Fluorescent Protein Leads to High Contrast Photoactivation". *J. Phys. Chem. Lett., 8 (13),* 2862-2868 (2017).

5.2: Introduction

Current techniques for superresolution imaging in live cells require fluorescent proteins that can be photoconverted between at least two spectrally-distinct states. One common route to photoconversion is called photoactivation, which refers to the conversion of an initially dark state to a brightly emissive state upon exposure to a certain wavelength of light. Perhaps the most famous example of this is photoactivatable GFP (PA-GFP), which, under 488 nm illumination, can be irreversibly converted from a dark to a bright state upon exposure to UV or blue light.[82] The mechanism for photoactivation in PA-GFP has been well-studied and is due to the decarboxylation of the carboxylate side chain of E222, which is thought to be initiated by charge transfer between E222 and the excited state of the fluorophore.[85] PA-GFP, and many other photoactivatable fluorescent proteins, have gained widespread use as photolabels for the collection of superresolution images in live cells.[97,99,100]

Here, we report the photoactivation abilities of superfolder GFP (sfGFP) and one of its point mutants. We replaced the threonine at position 203 (T203) in sfGFP with aspartic acid. The fluorophore that formed remained neutral between pH 5-10 indicating an apparent pK_a of at least 10. Exposure of this mutant to UV light allowed the

fluorophore to be converted to the negatively charged form, giving rise to a similar fluorescence contrast to that of PA-GFP. Mass spectrometry before and after exposure to UV light revealed a change in mass of 87 Da, which we attribute to the double decarboxylation of E222 and the newly inserted D203. This suggests that charge transfer from D203 to the excited state of the fluorophore results in decarboxylation of D203, presumably following a similar mechanism to the well-known phenomenon involving E222. The photoconversion process followed did not follow first-order kinetics, which may reflect the different charge transfer efficiencies between the fluorophore and E222 and D203. Non-linear power-dependence of the initial rate of photoconversion indicated that the double decarboxylation occurred via a multiphoton absorption process under 254 nm excitation. All together, these observations suggest that fluorescent proteins with novel photoconversion properties could be engineered by rationally inserting carboxylic acid residues near the fluorophore that can undergo charge transfer to the fluorophore and ultimately become decarboxylated.

5.3: Results and Discussion

The embedded fluorophore in the wild type form of green fluorescent protein (GFP) is formed from the residues S65-Y66-G67, which can be either neutral or negatively charged depending on the protonation of Y66. The neutral form (A state) absorbs maximally at 400 nm and is present in a ~6:1 ratio relative to the negative form (B state), which absorbs maximally at 490 nm.[62] Excitation of either of these states leads to green fluorescence with a high quantum yield (~0.6), either directly from the excited B state or from a deprotonated form of the excited A state after excited state proton transfer.[38,116] Although fluorescence is the dominant decay pathway of these excited states, several infrequent charge transfer pathways exist that lead to interesting

fluorophore chemistry, including bleaching, oxidative reddening, and decarboxylation of E222. Upon repeated illumination, these pathways can lead to the accumulation of various photoproducts.[117,118] For example, the accumulation of decarboxylated E222 in GFP causes an increase in the B state absorption at the expense of the A state. This is ultimately due to the change in fluorophore pK_a that accompanies the charge neutralization of E222. Optimization of the residues surrounding the fluorophore led to the development of PA-GFP, which exhibits a high fluorescence contrast upon accumulation of decarboxylated E222 and the concomitant population shift from neutral to charged fluorophore.[82]

The focus of our interest here centers around the decarboxylation of E222, which is shown mechanistically in Figure 5.1, following the Kolbe-type mechanism proposed by van Thor, et. al.[85] Following the generation of an excited state, there is a small but nonzero probability that E222 can donate its delocalized electron to the electron-deficient imidizolanone ring of the fluorophore. This is followed by fast transfer of a proton and an electron back to the E222 radical that ultimately results in decarboxylation of E222 and deprotonation of the fluorophore. While no intermediate species predicted by this mechanism have been directly observed, considerable experimental and theoretical evidence support this mechanism. van Thor, et. al. reported crystal structures of wild type GFP before and after exposure to UV light and observed the loss of electron density in the carboxylate side chain of E222. Mass spectrometry also revealed a mass reduction of 44 Da, which indicated the formation of one CO_2 equivalent according to the mechanism.[85] Bell, et. al. showed that the efficiency of the decarboxylation decreases with increasing illumination wavelength in the order of 254 nm > 280 nm > 476 nm.[93] They proposed that increased efficiencies in the UV range are due to the population of higher excited states that are stronger oxidizers and thus can accept an electron from

E222 more efficiently. In both studies by van Thor, et. al. and Bell, et. al., the authors observed that the rate of photoconversion followed first order kinetics, and the independent studies generally agreed upon the observed rate constant. Grigorenko, et. al. performed electronic structure calculations that showed that charge transfer states between E222 and the excited fluorophore could feasibly be populated by several pathways.[118] Additionally, the authors showed that one of the key residues in PA-GFP (H203) serves to lower the energy of one possible charge transfer state, which supports the hypothesis that the decarboxylation process is initiated by charge transfer between E222 and the fluorophore.

In contrast to the S65-Y66-G67 fluorophore (hereafter SYG), the one formed from T65-Y66-G67 (Figures 5.2A and 5.2B; hereafter TYG) has several interesting properties. First, the pK_a of Y66 in the TYG fluorophore is lowered such that the B state absorption dominates at physiological pH. This has largely to do with the extra methyl group on T65, which forces E222 to donate a hydrogen bond to the –OH side chain of T65, ultimately allowing E222 to be neutral even at high pH.[77,91] The neutral E222 allows the negatively charged form of the fluorophore to be favored, compared to the mutants with S65 where E222 can be negative and destabilize the B state. Additionally, the TYG fluorophore matures more quickly, has a higher extinction coefficient, and is more photostable which has earned mutants containing the TYG fluorophore the title of 'enhanced' GFPs due to their improved use as fluorescence tags for microscopy experiments.[83]

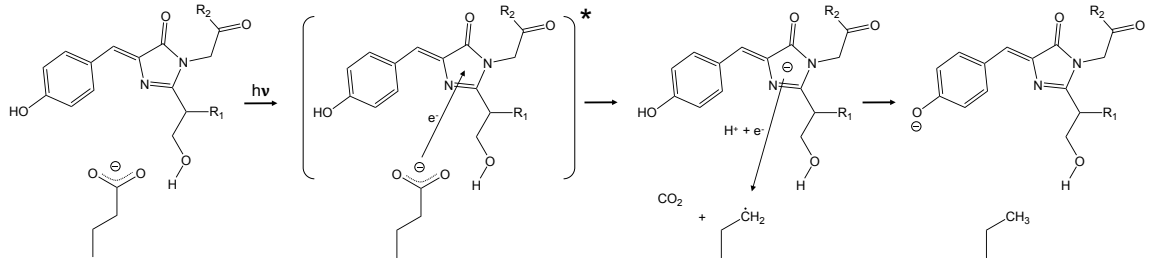

Figure 5.1: Mechanism of the Kolbe-type decarboxylation of E222 as it occurs in fluorescent proteins with the S65-Y66-G67 fluorophore. Upon the generation of an excited A state (represented by the asterisk), E222 transfers charge to the electron-deficient imidazolinone ring of the fluorophore, which results in the generation of a radical intermediate and the release of CO_2. The alkyl radical is quickly quenched, presumably by back transfer of an electron and a proton of the fluorophore, resulting in the stable decarboxylated E222 and a deprotonation of the phenolic oxygen of the fluorophore. Redrawn from ref. 85.

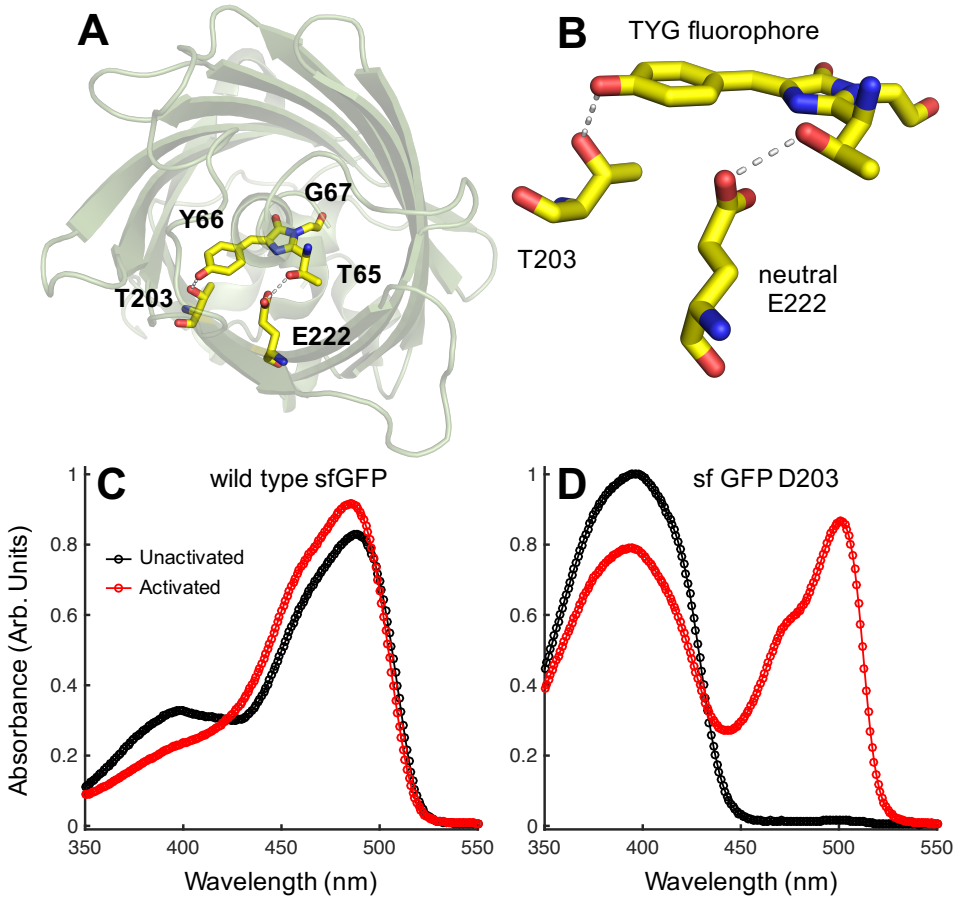

Figure 5.2: (A) Top-down view of the crystal structure of superfolder GFP (2B3P) showing the embedded T65-Y66-G67 fluorophore and residues E222 and T203. (B) A closer view of the fluorophore in superfolder GFP. Interactions between T203 and the phenolate of Y66 are shown, as well as the interaction between E222 and T65 which causes E222 to remain neutral (see main text). (C) The absorption spectrum of wild type superfolder GFP before and after 10 minutes of exposure to UV light. (D) The absorption spectrum of superfolder GFP T203D before and after 10 minutes of exposure to UV light.

While the TYG fluorophore offers spectroscopic benefits, it is inherently incompatible with the high contrast photoactivation that is observed in PA-GFP due to its increased B state absorbance. Additionally, if E222 is indeed neutral in GFP mutants with T65, then decarboxylation of this group would not represent as drastic of an electrostatic change and would not be expected to shift the fluorophore equilibrium significantly. Furthermore, the reduction potential of neutral glutamic acid is much lower than that of the negatively-charged carboxylate, which means that neutral E222 will be less likely to undergo the initial electron transfer process that is necessary for decarboxylation. To illustrate these points, we attempted to photoactivate wild type sfGFP, which contains the TYG fluorophore. Figure 5.2C shows the absorption spectrum of sfGFP before (black circles) and after (red circles) 10 minutes of exposure to 254 nm light with a power density of ~7 mW/cm^2. As expected, because of the dominant B state absorption before UV exposure, we observed only a minimal increase in the B state absorption (~10%) after 10 minutes of UV exposure.

We then became interested in making mutations that would allow the TYG fluorophore in sfGFP to be photoactivated with a similar contrast to that of PA-GFP, which contains the SYG fluorophore. Because one of the main differences between the SYG and TYG fluorophores is the charge on E222, we reasoned that a negatively charged residue near TYG would be necessary to force the equilibrium toward the A state such that its absorption spectrum resembled that of SYG. To do this, we replaced T203 with D203 because of the proximity of position 203 to Y66 (Figures 5.2A and 5.2B), and because of the well-known influence of the position 203 side chain on the fluorophore pK_a.[84,92] Serendipitously, D203 alone was enough to force the TYG equilibrium entirely to the A state (Figure 5.2D; black circles). To test whether this D203 mutant could be photoconverted, we exposed it to UV light using the same conditions as before. The red

circles in Figure 5.2D show the D203 mutant spectrum after 10 minutes of UV exposure. Interestingly, we observed a very high contrast in the B state absorption, similar to that of PA-GFP. We note that Patterson and Lippincott-Schwartz made the D203 mutation in their development of PA-GFP, but observed no absorption of the resulting protein between 350-550 nm for undetermined reasons.[96] This was likely due to improper folding or the absence of fluorophore formation, which appears to be overcome here by the use of sfGFP, which has robust folding properties and a high tolerance for random mutations.

To confirm that photoconversion of the TYG fluorophore in the D203 mutant occurred following a similar mechanism to the one outlined in Figure 5.1, we performed electrospray ionization mass spectrometry using a quadrupole mass analyzer to determine the difference in mass before and after the exposure to UV light. Similar to previous reports, we expected to see a mass loss of 44 Da which would indicate the formation of one CO_2 equivalent according to the above mechanism. Figure 5.3A shows the resulting mass spectrum of the D203 mutant before UV exposure. The most abundant peak, with a mass of 27887.5 Da, corresponds to the full protein chain including a hexa-histidine affinity tag. The smaller mass corresponds to the same chain minus the N-terminal methionine residue, which is commonly lost during ionization. Figure 5.3B shows the resulting mass spectrum of the D203 mutant after it was exposed to UV light as described above. Interestingly, the most abundant mass in this spectrum (27800.0 Da) corresponds to the pre-exposure mass from Figure 3B minus 87.5 Da. There is also a mass loss of 87.3 Da for the protein chain that lacks the N-terminal methionine. These mass changes are strongly indicative of the loss of two CO_2 groups, which can only be explained by the decarboxylation of both E222 and the newly inserted D203. Further inspection of Figure 3B also shows several other masses that were not present in Figure 5.3A, which

correspond to the singly decarboxylated species and a small abundance of protein that did not decarboxylate at all.

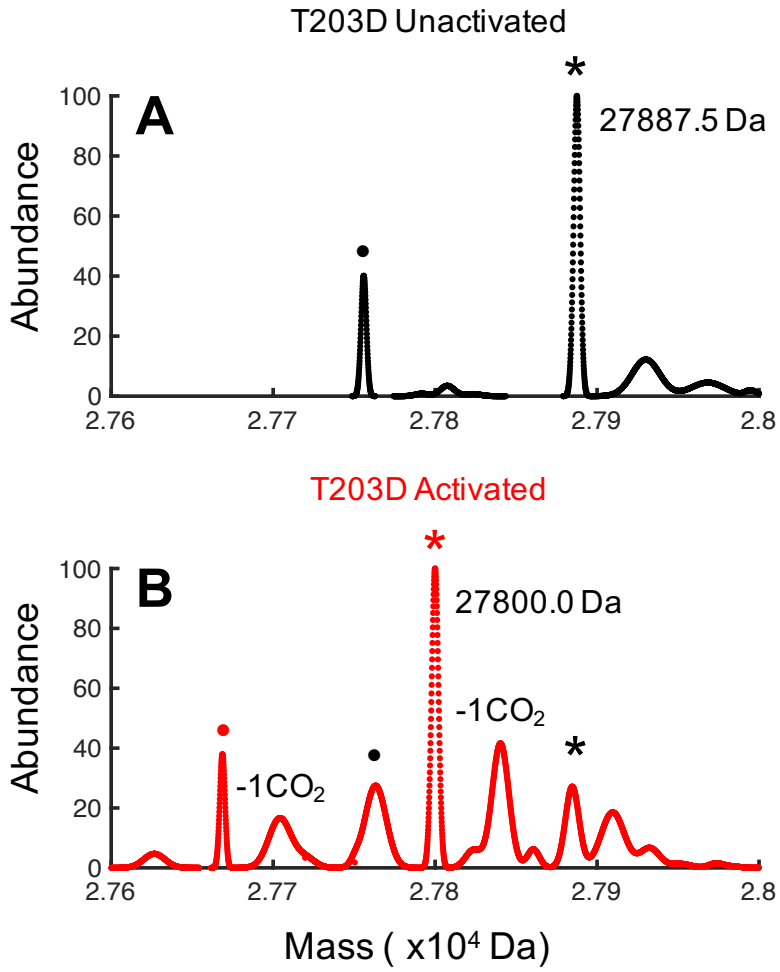

Figure 5.3: Exposure to UV light leads to a mass reduction of 87.5 Da. (A) ESI-MS of superfolder GFP T203D before UV exposure shows two main species, which correspond to the full protein chain (*) and a chain missing the N-terminal methionine (•). The dominant species is 27887.5 Da. (B) After UV exposure, both species from (A) shift in mass by ~87.5 Da. Two new species appear after UV exposure which correspond to singly decarboxylated fragments (indicated by -1 CO_2). Additionally, there is still abundance of proteins that do not undergo decarboxylation (indicated by the same labels used in (A)).

This finding is astonishing, because it suggests that D203 can also donate charge to the excited fluorophore to initiate a decarboxylation similar to the one outlined in Figure 5.1. If D203 can also initiate the charge transfer, then the site of initial charge transfer could be a tunable parameter in the design of fluorescent proteins with increased photoconversion efficiencies. To assess the utility of the D203 mutant as a high contrast fluorescence agent, we recorded the fluorescence spectrum under 488 nm excitation before and after the UV exposure. This is shown in Figure 5.4A, where we observed a ~95-fold increase in the integrated fluorescence after 10 minutes of UV exposure. This is similar to the activation efficiency that occurs with PA-GFP,[82] although it remains to be seen if this D203 mutant would retain such high contrast when imaged in live cells. However, one can expect that the desirable properties of sfGFP were not completely suppressed by the single point mutation we made to place aspartic acid at position 203. In this regard, we have constructed a photoactivatable variant of sfGFP, which could offer improvements over PA-GFP in studies where the robust folding properties of sfGFP are required. Notably, Berlin, et. al. reported a similar mutant in 2015, where they made the mutations of T203H, T65S, and F64L in sfGFP to revert it back to an analog of PA-GFP.[98] However, to our knowledge, the mutant in this study is the first mutant to contain the desirable TYG fluorophore and exhibit high contrast photoactivation.

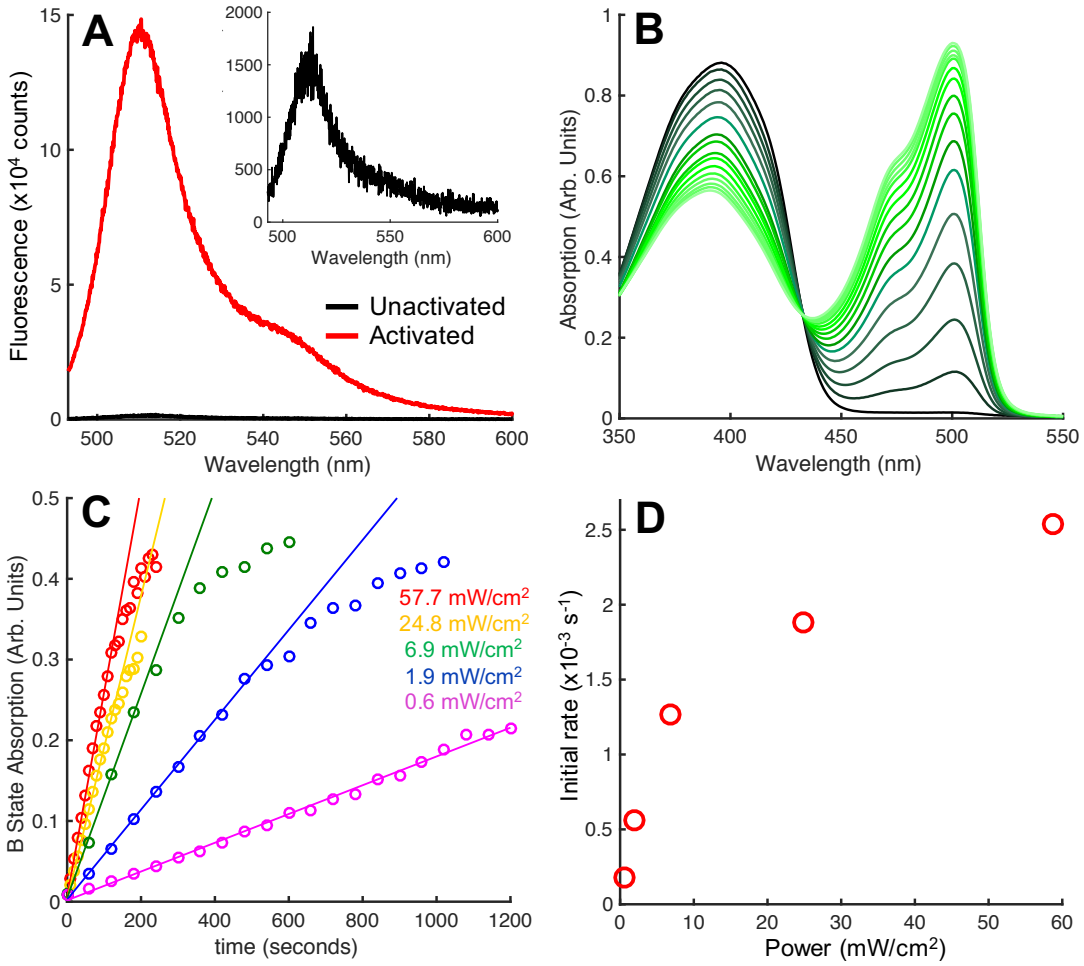

Figure 5.4: Spectroscopic and kinetic properties of photoconversion. (A) Fluorescence of the T203D mutant before (black) and after (red) 10 minutes of UV exposure indicating a ~95-fold increase in the fluorescence under 488 nm excitation due to the photoconversion. (B) Absorption spectra of the TYG fluorophore in the T203D mutant as a function of UV irradiation time, plotted at 1 minute intervals for a total of 15 minutes. The black spectrum was taken before irradiation and the greener spectra correspond to later time points. (C) The B state absorption as a function of time for several different UV irradiation powers. Best-fit lines are drawn over the time course that corresponds to the linear portion of the curve to estimate the initial rates. (D) The initial rates of photoconversion from (C) plotted against irradiation power.

Regardless of its use as a fluorescence label in live cells, this D203 mutant of sfGFP offers significant insight into photo-induced chemistry in fluorescent proteins. If charge transfer (CT) can occur from either E222 or D203 to the excited state of the fluorophore, then it might be expected that the CT processes occur with different efficiencies. It has been proposed to improve CT efficiencies by mutating residues that can lower the energies of CT states or improve the oxidizing power of the fluorophore excited state.[93] However, the possibility of rationally moving the charge donor site closer to the acceptor to increase CT has so far not been suggested. Grigorenko, et. al. showed that the residue at position 203 can modulate the energies of the proposed charge transfer states, which makes populating them more efficient.[118] However, in the case of the D203 mutant, where charge transfer might actually occur from position 203, the more relevant factor in determining the efficiency of charge transfer might simply be the distance between the donor and acceptor. Stark spectroscopy has revealed that the imidazolinone ring is electron deficient in the A^* state,[39] which supports its role as the initial electron acceptor as drawn in Figure 5.1. If this holds true in the D203 mutant, then it should be the case that CT transfer from D203 is less efficient than CT from E222 due to its increased distance from the imidizolanone ring of the fluorophore (0.8 vs. 0.6 nm). As such, we expect that the mechanism for this double decarboxylation process might not obey first order kinetics if there are indeed two parallel reactions that take place with different efficiencies.

To measure the kinetics of the decarboxylation reaction, we irradiated the D203 mutant with the same power of UV light as before and recorded the absorption spectrum of the TYG fluorophore at regular intervals over the course of 15 minutes. The resulting spectra are shown in Figure 5.4B. We observed a clean isosbestic point at 432 nm, which indicated that the UV exposure induced a photoconversion between only two states. 15

minutes of exposure at this irradiation power caused conversion from a fluorophore population that existed almost entirely in the A state to a fluorophore population with a ~2:3 ratio of A to B state. Plotting the B state absorption as a function of the irradiation time (Figure 5.4C, green data points) for this photoconversion yielded a curve that was not well-described by first order kinetics, which is in contrast to the kinetics that have been observed for photoconversion reactions in GFP and PA-GFP. However, as we mentioned above, the presence of two CT processes in the same system occurring with different efficiencies might undoubteldly lead to a complex rate equation for this double decarboxylation. Preliminary attempts to quantify the kinetics suggest that the reaction might obey second-order or even quadratic kinetics. Further experiments are underway to build a fully descriptive kinetic model for this process, under the assumption that there are two parallel Kolbe-type mechanisms at play in this double decarboxylation.

Another interesting observation of this double decarboxylation process is the power dependence of the photoconversion rate. In Figure 5.4C, we show the kinetics of the photoconversion at different irradiation powers between 0.6 – 57.7 mW/cm^2. In all cases, the growth of the B state absorption as a function of time could not be well-described by a first-order kinetic process, which further supports the complex kinetics of the double decarboxylation. The best-fit lines drawn in Figure 4C are linear fits over the initial linear range of each trace, which we used to estimate the initial velocities of the reaction at different powers. Figure 5.4D shows the initial rates plotted against the corresponding irradiation powers. We clearly observed a non-linear power dependence of the rates over this range, which is indicative of a multiphoton process. This is in contrast to previously observed single decarboxylation events, which display linear power-dependence with 254 nm irradiation. If the double decarboxylation is indeed a two-

photon absorption process, there could be potential use for mutants like this one in two-photon microscopy.

Altogether, these observations on the photoconversion properties of sfGFP mutant T203D lead to several questions regarding the Kolbe-type decarboxylation mechanism that has been proposed to occur in many fluorescent proteins. First and foremost, does the decarboxylation of D203 occur via a similar mechanism to the one outlined in Figure 5.1? If so, then it might be the case that CT to the excited state of the fluorophore can be tuned to achieve more efficient photoconversion. Along these lines, assuming the mechanism of decarboxylation of D203 is similar to that of E222, then it should be the case that the CT efficiency between D203 and A^* is different than that between E222 and A^*. Moreover, given the proposal that the imidazolinone ring of the fluorophore is the initial electron acceptor, then it should be expected for CT to be more efficient from E222 compared to D203, because E222 is closer to the imidazolinone ring (~0.6 vs. 0.8 nm). However, because E222 in this mutant is likely to be neutral, it is expected that the CT efficiency from E222 should be relatively low because of the low concentration of negative carboxylate. Future studies of the kinetics of this decarboxylation should focus on elucidating the pK_a of both E222 and D203, because those values ultimately determine the concentrations of negative carboxylates, and thus the reactive species. Overall, we think that mutations of this nature could provide a tunable parameter in the design of fluorescent proteins with new properties which has so far been unexplored.

Chapter 6: Future Directions

6.1: Introduction

The final chapter of this dissertation will focus on three separate observations which have arisen throughout the course of the studies presented in Chapters 1-5. They each represent interesting findings that could provide fruitful future work upon their continuation. In section 6.2, we report the temperature dependent vibrational frequencies of *p*CNF at 9 positions throughout the sequence of GFP, similar to what was reported in Chapter 3. We show that the magnitude of the temperature dependence of these nitrile spectra is related to the solvent accessible surface area of the nitrile probe, which suggests that the temperature dependence could report quantitatively on hydrogen bonding environment. This is similar to a recent report by Adhikary, et. al, but their analysis was limited to only a qualitative description of the hydration status of nitrile probes.[113] We attempt to use these spectra to accurately quantify the total electric field experienced by nitrile probes, even in the background of hydrogen bonding.

Section 6.3 details a series of observations that have been made for the GFP mutant containing *p*CNF at position 145 and cysteine at position 203. This mutant was used in the Stark effect – pK_a comparison in Chapter 4, and here we focus more closely on interesting spectroscopic features of this mutant. Specifically, *p*CNF 145 appears to be sensitive to the protonation state of C203. pH-dependent nitrile absorption spectra suggest that C203 titrates with a pK_a of ~7. The pK_a of C203 on its own is not particularly interesting, but this represents an unusual method for measuring pK_a in a protein.

Finally, in section 6.4 we report the fluorescence lifetimes of nearly all of the GFP mutants that have been presented in this dissertation. The factors that contribute to differences in excited state lifetimes are of considerable interest from both practical and fundamental viewpoints. On their own, the lifetime measurements do not reveal any

significant insight about GFP. However, because of our choice of mutants containing the TYG fluorophore, and the insertion of nitrile VSE probes near the fluorophore, each lifetime measurement is coupled with site-specific electric field measurements as presented in Chapters 3 and 4. Our data suggest that changes in fluorescence lifetimes can be understood in terms of changes in electric fields measured from the vibrational and electronic probes. This is contrary to the long held belief of the fluorescent protein community, that fluorescence lifetimes are largely dominated by steric interactions which either promote or reduce non-radiative relaxation. Whatever the case may be, the GFP system we have developed, with site-specific electric field measurement capabilities, could provide an interesting experimental viewpoint into this rapidly evolving topic.

6.2: Quantifying the Hydrogen Bonding to Nitrile Probes

So far in this dissertation, we have reported the use of nitrile vibrational frequencies to report solely on *changes* in electric field, relative to a well-defined reference, via Equation 1.2. However, in principle, Equation 1.2 could be used to relate absolute frequencies to absolute electric fields. The problem in doing this with measured nitrile frequencies arises from the ability of the nitrile to accept hydrogen bonds, which causes deviations from the Stark effect model due to quantum mechanical effects associated with the hydrogen bond. This has led to suggestions that nitrile probes should only be used in environments where they can be free from hydrogen bonds, or to assess whether a particular location is solvent-exposed.[72,113] While one can certainly imagine the use of a nitrile probe in the latter case, it is difficult to imagine many biological scenarios where a nitrile probe might be free from hydrogen bonds entirely. In fact, it seems likely that a nitrile probe placed into a water-free environment such as the hydrophobic interior of a protein will simply accept a hydrogen bond from another donor besides water. This

remains to be rigorously tested, but the sentiment still remains that there are probably very few interesting biological environments where a nitrile probe could be entirely free from hydrogen bonds, aside from in simple, aprotic solvents. As such, if we are to use nitrile probes to measure electric fields in hydrogen bonding environments, then a rigorous quantification of the effect of hydrogen bonding on nitrile vibrational frequencies is desired.

Cho, et. al. performed quantum mechanics calculations on small water clusters containing methyl thiocyanate and acetonitrile in order to assess the frequency changes that are accompanied by hydrogen bonding with water.[71] Their results showed that the nitrile frequencies varied strongly with the orientation between the hydrogen bond donor and the nitrile bond vector. Boxer, et. al. showed that, compared to an independent NMR parameter, nitriles that engaged in hydrogen bonds with water all had frequencies that were offset by a constant amount from the NMR-IR comparison.[73] An unfortunate result of this elegant study is that the constant offset of 10 cm^{-1} that was observed has been taken by many to be the constant value that should be applied to *all* hydrogen bonded nitriles in all hydrogen bonding environments. An interesting future study would be to extend the comparison to hydrogen bonding environments other than just small molecules in water or solvent-exposed nitrile probes.

To understand the effects of hydrogen bonds from different hydrogen bond donors, we have recorded FTIR spectra of methyl thiocyanate in a wide range of organic solvents. This is shown in Figure 6.1, where the peak center frequencies are plotted against π^*, which is an empirically derived parameter that attempts to capture the polarity and polarizability of different solvents.

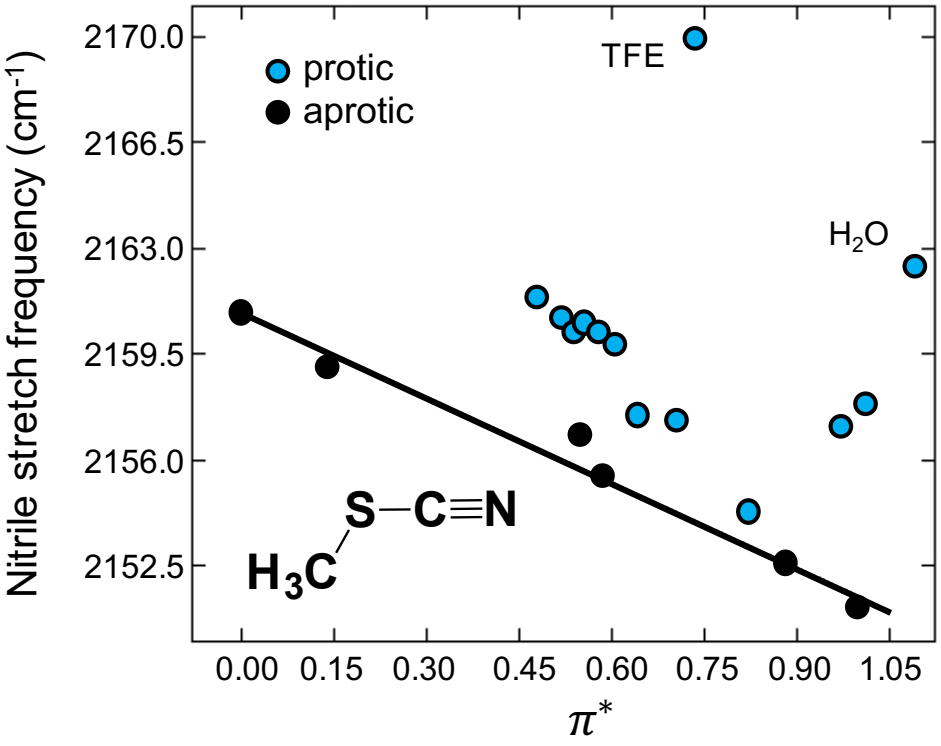

Figure 6.1: Nitrile frequencies of methyl thiocyanate plotted against solvent polarity. The chemical structure of methyl thiocyanate is shown in the bottom left. π^* represents the solvent polarity such that larger values of π^* indicate a more polar solvent. Protic solvents are shown in blue and aprotic solvents are shown in black. The best-fit line shown is drawn for the aprotic data points only. The data points corresponding to TFE (trifluoroethanol) and water are specifically labelled. Error bars are omitted for clarity, but the standard deviation of at least three replicate measurements was less than 0.1 cm^{-1} in all cases.

In Figure 6.1, the best-fit line is drawn for the aprotic solvents only. We observed a strong linear correlation between nitrile frequency and the polarity of aprotic solvents. This is similar to previous observations by other researchers who have shown that simple solvation models can account for nitrile frequencies in non-hydrogen bonding environments.[63,119] However, we clearly observed no unifying trends between frequency and polarity for the generally blue-shifting hydrogen bonding environments. Similar studies have observed exactly the same effect, but have focused solely on the hydrogen bonding effects of water alone. While hydrogen bonds from water are certainly the most prevalent in any studies which use nitrile probes in biological systems, their effects are clearly not representative of all hydrogen bonds. As such, it would be desirable to have an understanding of the determinants of nitrile frequencies in the presence of all types of hydrogen bonds.

A few attempts have been made to understand the effect of hydrogen bonds on nitrile frequencies in scenarios that are not simply a nitrile probe in a bulk-water environment.[73] However, these studies have largely focused on varying the ratios of mixed co-solvents with water. Bagchi, et. al. have shown that nitrile frequencies vary linearly with water content in water-DMSO mixtures, for example.[120,121] Studies like these are often used to say that nitrile frequencies can be related to hydrogen bonding environment in a wide-range of hydrogen bonding scenarios. While these studies represent a step in the right direction, it is unclear that adding co-solvents should necessarily be a good stand-in for the different hydrogen bonding environments that one might find in an enzyme active site or at a protein-protein interface, for example. It seems more likely that the observed linear increases in nitrile frequency that accompany increases in the DMSO:water ratio might simply be due to linear increases in the amount of hydrogen bonding, without any drastic changes to the geometry of the interaction.

Moreover, studies like these ignore the fact that many hydrogen bond donors besides water exist in biological systems.

One promising method that could allow for the assessment of the hydrogen bonding environments of nitrile probes is the measurement of their temperature dependence. Adhikary, et. al. first reported this in 2015 and showed that the magnitude of the frequency-temperature line slope (FTLS) is related to a qualitative expectation of the hydrogen bonding environment based on the observation of crystal structures.[113] They reported that nitrile probes inserted at solvent-exposed locations had larger FTLSs than nitriles inserted at buried locations. Qualitatively, this makes sense from the viewpoint that hydrogen bonds tend to broaden and blue shift the frequencies of nitriles, relative to their unbound values (see Figure 6.1, for example). Thus, as temperature increases and hydrogen bonds occur less frequently, there should be a corresponding red shift and narrowing of the nitrile spectrum.

In order to connect these qualitative observations to quantitative information about the hydrogen bonding environments of nitrile probes, we have inserted *p*CNF throughout the structure of GFP and measured the FTLS for all of these nitrile probes. Figure 6.2 shows crystal structure 2B3P with all of the *p*CNF insertion locations highlighted. Of the locations, two are expected to be mostly solvent exposed (149, 212), three are slightly buried (114, 143, and 218), and three should have their side chains oriented entirely into the beta-barrel of GFP (92, 145, 165). Figure 6.3 shows the temperature dependence of the *p*CNF frequencies at each of these eight locations. The data for *p*CNF in water is shown for reference. We observed relatively small FTLS values for the two buried nitriles at positions 145 and 165, which indicated that *p*CNF at those positions are only weakly engaged in hydrogen bonds. Conversely, for the well-solvent exposed location of 212 on the extended loop (see Figure 6.2) we observed a very

large temperature dependence, which suggests that *p*CNF at that position is strongly engaged in hydrogen bonds. Interestingly, the nitrile with the highest FTLS was *p*CNF 92, whose side chain is pointed entirely into the barrel in crystal structure 2B3P. Because of the abundance of confined waters within the interior of GFP, we hypothesize that the nitrile at position 92 may be constrained to an orientation where it almost always accepts a strong hydrogen bond, from either a constrained water molecule or a protein hydrogen bond donor.

Overall, we observed that *p*CNF probes at each of these eight locations gave distinct FTLS values, which suggests they all experienced different hydrogen bonding environments. Future work on this project will relate to quantifying the hydrogen bonding of the *p*CNF probes in atomistic dynamics simulations. We aim to link the experimentally-measured FTLS values to the molecular-level details in the simulations. Doing so would provide an experimental route towards quantifying the hydrogen bonding environments of nitrile probes, which would enhance their overall utility as probes of local environment.

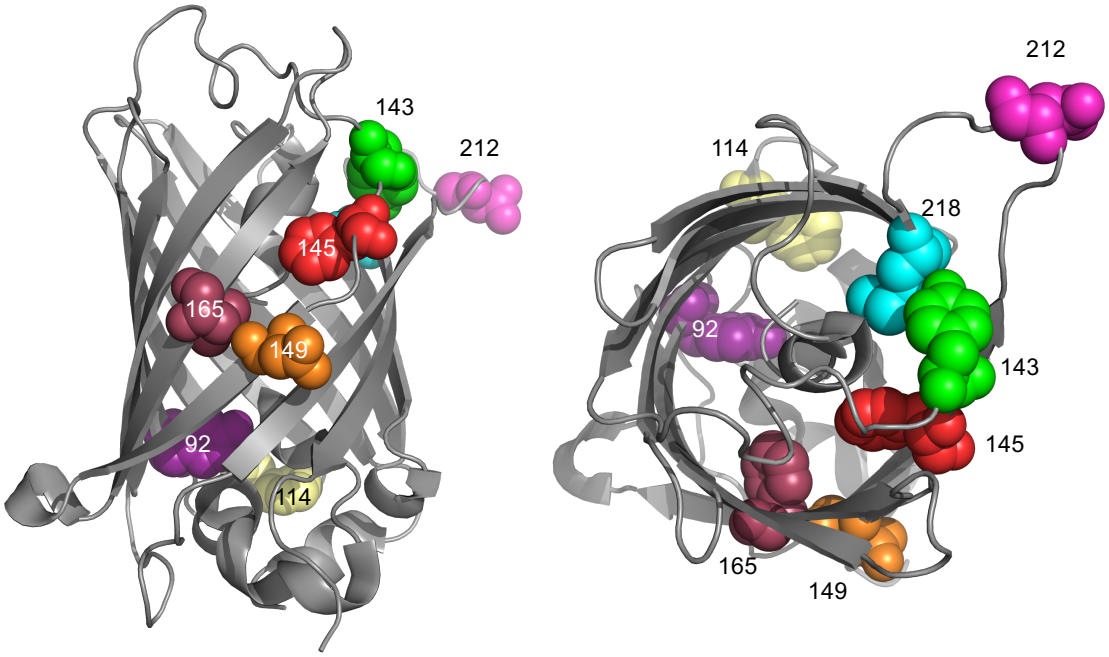

Figure 6.2: Two views of crystal structure 2B3P showing the eight amino acid positions (colored spheres) that were replaced with *p*CNF. The number next to each position represents the position of that residue in the primary sequence of GFP.

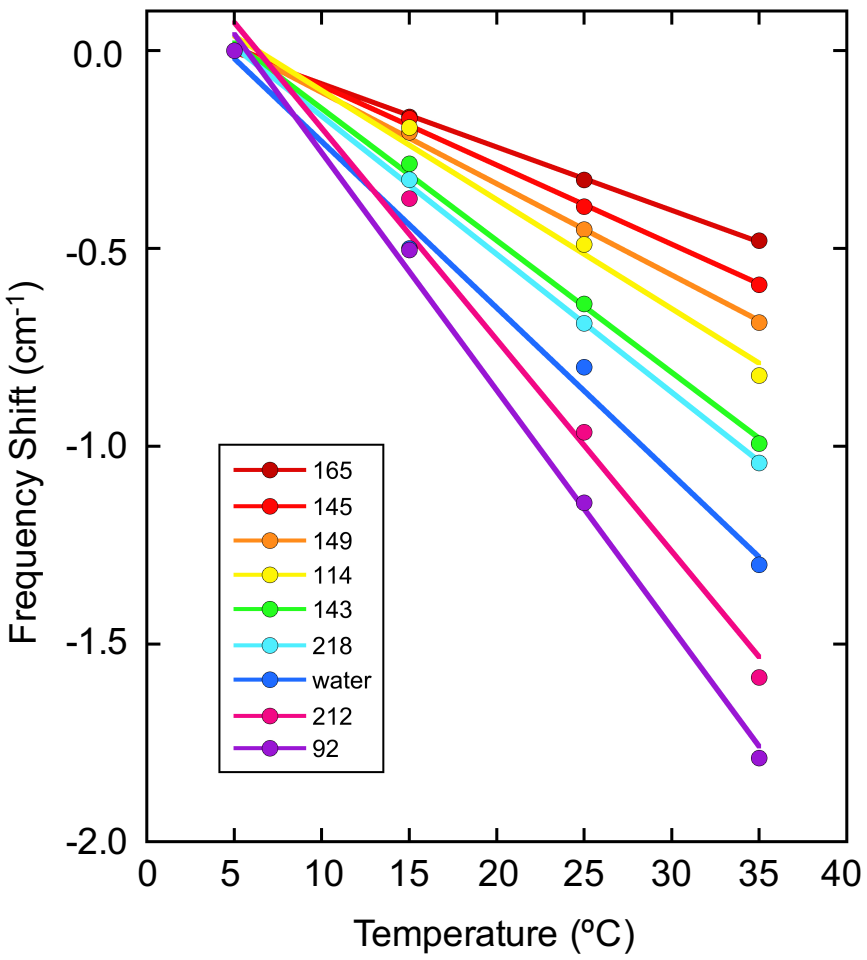

Figure 6.3: The temperature dependence of *p*CNF spectra from the eight locations shown in Figure 6.2. The y-axis is the frequency shift relative to the value at 5 °C and the x-axis is temperature. Different colors represent the different *p*CNF positions based on the key and correspond to the color scheme used in Figure 6.2. Error bars are omitted for clarity, but the standard deviations of at least three measurements were less than 0.1 cm^{-1} for all data points.

6.3: Measuring the pK_a of Cysteine 203 with *p*CNF 145

The GFP mutant containing cysteine at position 203 and *p*CNF at position 145 (which we refer to as FPC 145) exhibits interesting UV-Vis and infrared spectra. Figure 6.4 shows the UV-Vis spectra of GFP mutants containing cysteine at position 203 and no nitrile (black), *p*CNF 145 (blue), or *p*CNF 165 (red). All of the spectra contain the two characteristic features of the A and B state, present in different ratios due to their different pK_a values as detailed in Chapter 4. Additionally, the two nitrile-containing mutants have a third feature around 470 nm. This feature is seen in the absorption spectra of many GFP fluorophore (see Figure 4.4, for example) and is usually attributed to an alternative environment of the B state. However, the nature of this environment is not known and is not well-studied. Interestingly, the absorption spectrum of the FPC 145 mutant in Figure 6.4 shows an enhancement of this feature, as well as a strong red shift in the A state absorption. A systematic study of this mutant could help reveal the nature of the ~470 nm feature.

One question that follows naturally from the observation of this enhanced feature in Figure 6.4 is: what does the vibrational spectrum of *p*CNF 145 look like? We have already reported the FTIR spectrum of this nitrile-containing mutant in Chapter 4. Figure 4.4D showed that, of all the nitrile vibrational spectra, FPC 145 was the only one that could not be well-described by a single Gaussian function. Because of the proximity of position 145 to the fluorophore, and because of the interesting features in the fluorophore spectrum, we wanted to see if the *p*CNF 145 spectra in this mutant responded to pH. Figure 6.5 shows the *p*CNF 145 spectra of this mutant at a range of pH values between 6-8.5. Over this range, there appears to be at least two features whose relative ratios can be modulated by pH. Based on the ratio of these features as a function of the solution pH, the nitrile spectra appear to be reporting on a proton equilibrium that has an apparent pK_a

of roughly 7. In Chapter 4, we reported that the fluorophore pK_a of this mutant is 7.86, so there are a couple of interesting implications of the vibrational spectra shown in Figure 6.5. First, the pH-dependent nitrile spectra could be reporting on the equilibrium of the fluorophore. In this case, the pK_a values of ~7 and 7.86 are quite different and so would represent a difference in the pK_a value depending on the method used to measure the pK_a. This is potentially very problematic. Another more likely scenario is that the *p*CNF 145 is relatively insensitive to the A to B state equilibrium, as has been the case for the other GFP mutants with these nitrile probes. This would suggest that *p*CNF is sensitive to the titration of C203, and that the apparent pK_a from these vibrational spectra should be assigned to C203. Additionally, the presence of C203 with an apparent pK_a of ~7 means that this residue is present in both neutral and negatively charged forms at physiological pH, might be able to explain the appearance of the enhanced feature at 470 nm in the absorption of the GFP fluorophore. Future studies should be aimed at molecular modeling of these GFP mutants with C203, as well as the independent determination of the pK_a of C203 through other methods such as NMR.

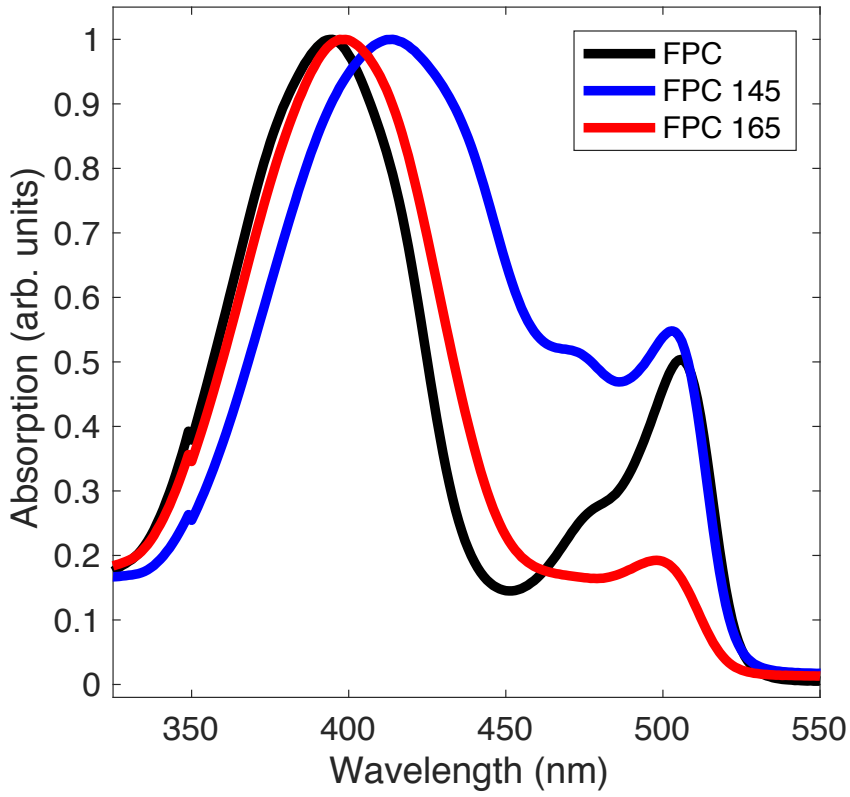

Figure 6.4: GFP fluorophore absorption of mutants containing cysteine at position 203 and either no nitrile or a *p*CNF at position 145 or 165.

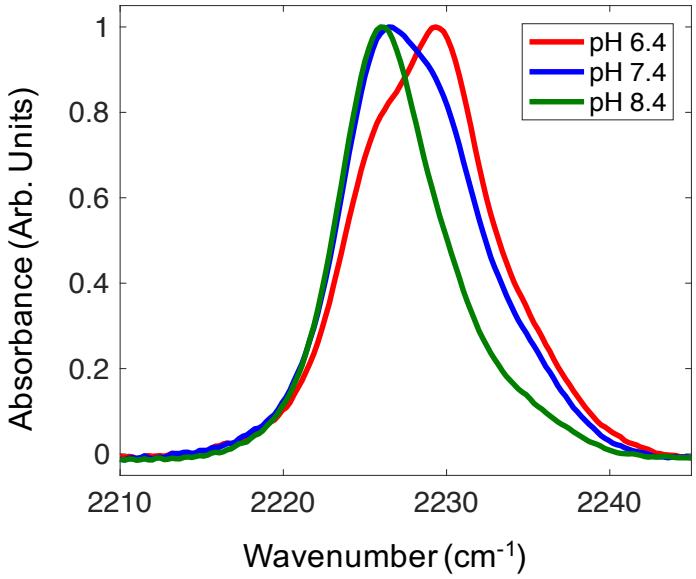

Figure 6.5: Nitrile FTIR spectra of the GFP mutant containing cysteine at position 203 with *p*CNF 145, measured between pH 6.4 – 8.4 as indicated by the color key.

6.4: Relating Fluorescence Lifetimes to Protein Electric Fields

The photophysical properties of the GFP fluorophore depend very sensitively on the local electrostatic environment in the interior of the barrel structure. Because we have put forth such a great effort towards characterizing the effect of systematic changes to position 203 on the electrostatic environment of the fluorophore, we are in a unique position to be able to understand how these changes in electrostatic environment lead to changes in various photophysical properties. Here, we report measurements of the fluorescence lifetimes of GFP, in the hope that these measurements will stimulate future studies to elucidate the determinants of excited state lifetimes in GFP. Figure 6.6 shows the results of time correlated single photon counting measurements for each of the mutants that we studied in Chapter 3. Figures 6.6A-C are the measured decays for mutants no nitrile probe (A), *p*CNF 145 (B), and *p*CNF 165 (C). The decays were fit to single exponentials to estimate the lifetimes, which are summarized in Figure 6.6D. Error bars represent the instrument resolution, which was much larger than the standard deviations of at least three measurements for all mutants. We observed lifetimes between 2.3-3.5 ns. Similar to what is known about the GFP fluorophore, we observed the long lifetimes when aromatic residues were at position 203 and short lifetimes when small, polar residues were at position 203.[122] With the exception of S203 and H203, the insertion of either nitrile probe tended to decrease the observed lifetime, with *p*CNF 165 usually having a larger effect than *p*CNF 145.

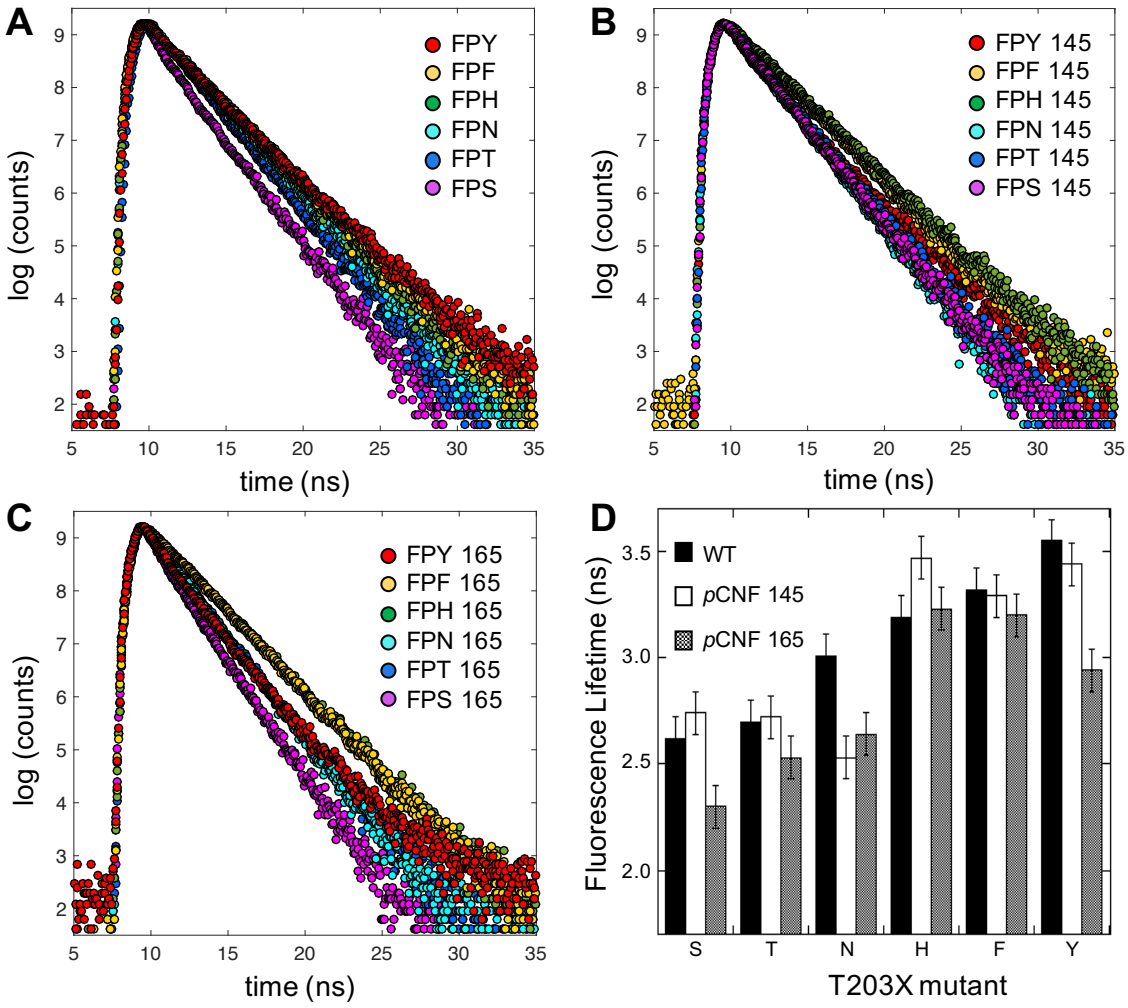

Figure 6.6: Time-correlated single photon counting measurements of GFP mutants containing no nitrile probe (A), *p*CNF 145 (B), or *p*CNF 165 (C). Different colors indicate a different amino acid side chain at position 203. Panel (D) shows the results of single exponential fits to the decays from (A) through (C), where the letter on the x-axis denotes the identity of the amino acid at position 203. Error bars represent the instrument resolution of 100 ps. In the keys, 'FP' stands for 'fluorescent protein' and the letter that follows designates the one letter amino acid code of the residue at position 203.

Figure 6.7 shows the changes in fluorescence lifetime for all pairs of position 203 mutants, plotted against the resulting electric field change inferred from the B state absorption energies via Equation 1.2. Circles represent the data for GFP mutants with no nitrile probe, and the mutants with *p*CNF 145 and *p*CNF 165 are represented with squares and diamonds, respectively. Remarkably, we observed a strong linear correlation with $r > 0.8$ for all comparisons. Slight variations in the slopes of the best-fit lines represent the slight tendency of the inserted *p*CNF probe to disrupt the lifetime-field correlation. Additionally, in all cases we observed similar trends to what we have described previously in Chapters 3 and 4, where specific types of chemical modifications cause similar changes in the fluorescence lifetimes (as indicated by the colors). These data suggest that an electrostatic mechanism could be at play in modulating the lifetime of the GFP fluorophore. Future directions should include mutations to position 203 other than those listed here, as well as mutations to other position near the GFP fluorophore which undoubtedly play important roles in determining these electrostatic effects in the GFP fluorophore cavity.

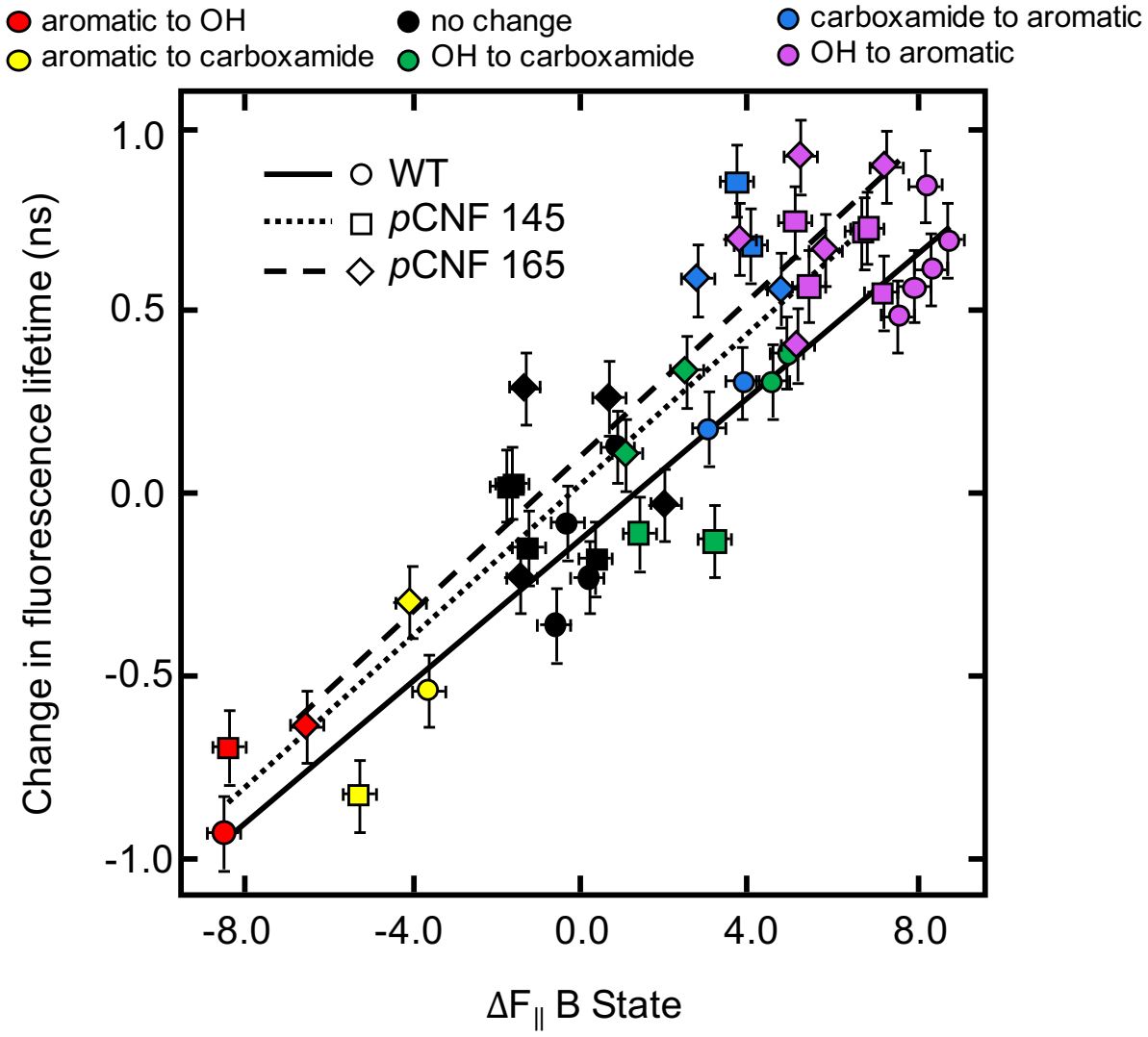

Figure 6.7: Changes in fluorescence lifetime plotted against changes in electric field measured from the B state absorption of the GFP fluorophore via Equation 1.2. Circles represent mutants with no nitrile probes, and squares and diamonds represent mutants with *p*CNF 145 and 165, respectively. The colors represent the type of mutation that was made a position 203 according to code shown above the plot. Error bars on both the x-axis and y-axis represent the instrument resolution and are not indicative of variability in the measurements, which was well below the manufacturer quoted instrument resolution.

References

(1) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106* (8), 3210–3235.

(2) Warshel, A. *Proc. Natl. Acad. Sci. USA* **1978**, *75* (11), 5250–5254.

(3) Honig, B.; Nicholls, A. *Science* **1995**, *268* (5214), 1144–1149.

(4) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A.; Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. **2000**, *289* (5481), 920–930.

(5) Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946–950.

(6) Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. *Proteins Struct. Funct. Genet.* **2002**, *48* (2), 388–403.

(7) Langsetmo, K.; Fuchs, J. A.; Woodward, C. *Biochemistry* **1991**, *30* (30), 7603–7609.

(8) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; Velu, P. D.; García-Moreno E., B. *Proc. Natl. Acad. Sci.* **2010**, *107* (37), 16096–16100.

(9) Bradbury, J. H.; Scheraga, H. A. *J. Am. Chem. Soc.* **1966**, *88* (18), 4240–4246.

(10) Markley, J. L. *Acc. Chem. Res.* **1975**, *545* (6), 70–80.

(11) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; García-Moreno E, B. *Biophys. J.* **2000**, *79* (3), 1610–1620.

(12) Chimenti, M. S.; Castañeda, C. A.; Majumdar, A.; García-Moreno, B. E. *J. Mol. Biol.* **2011**, *405* (2), 361–377.

(13) Inoue, M.; Yamada, H.; Yasukochi, T.; Kuroki, R.; Miki, T.; Horiuchi, T.; Imoto, T. *Biochemistry* **1992**, *31* (24), 5545–5553.

(14) Davoodi, J.; Wakarchuk, W. W.; Campbell, R. L.; Carey, P. R.; Surewicz, W. K. *Eur. J. Biochem.* **1995**, *232* (3), 839–843.

(15) McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Korner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35* (31), 9958–9966.

(16) Isom, D. G.; Cannon, B. R.; Castañeda, C. A.; Robinson, A.; García-Moreno, B.

Proc. Natl. Acad. Sci. U. S. A. **2008**, *105* (46), 17784–17788.

(17) Gibas, C. J.; Subramaniam, S. *Biophys. J.* **1996**, *71* (1), 138–147.

(18) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35* (24), 7819–7833.

(19) Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Recognit.* **2002**, *15* (6), 377–392.

(20) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins Struct. Funct. Genet.* **2005**, *61* (4), 704–721.

(21) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

(22) Witham, S.; Talley, K.; Wang, L.; Zhang, Z.; Sarkar, S.; Gao, D.; Yang, W.; Alexov, E. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3389–3399.

(23) Meyer, T.; Knapp, E. W. *J. Chem. Theory Comput.* **2015**, *11* (6), 2827–2840.

(24) Mehler, E. L.; Guarnieri, F. *Biophys. J.* **1999**, *77* (1), 3–22.

(25) Schutz, C. N.; Warshel, A. *Proteins Struct. Funct. Genet.* **2001**, *44* (4), 400–417.

(26) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta - Proteins Proteomics* **2006**, *1764* (11), 1647–1676.

(27) Nielsen, J. E.; Gunner, M. R.; García-Moreno E., B. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3249–3259.

(28) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 243–252.

(29) Warshel, A.; Dryga, A. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3469–3484.

(30) Alexov, E.; Mehler, E. L.; Baker, N.; M. Baptista, A.; Huang, Y.; Milletti, F.; Erik Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3260–3275.

(31) Li, L.; Li, C.; Zhang, Z.; Alexov, E. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

(32) Jr, R. D. G.; Kieslich, C. A.; Nichols, A.; Sausman, N. U.; Foronda, M.; Morikis, D. *Biopolymers* **2011**, *95* (11), 16–19.

(33) Getahun, Z.; Huang, C. Y.; Wang, T.; De León, B.; Degrado, W. F.; Gai, F. *J. Am.*

Chem. Soc. **2003**, *125*, 405–411.

(34) Lockhart, D. J.; Boxer, S. G. *Biochemistry* **1987**, *26* (3), 664–668.

(35) Lockhart, D. J.; Boxer, S. G. *Proc. Natl. Acad. Sci. USA* **1988**, *85* (1), 107–111.

(36) Lockhart, D. J.; Kirmaier, C.; Holten, D.; Boxer, S. G. *J. Phys. Chem.* **1990**, *94* (18), 6987–6995.

(37) Steffen, M. A.; Lao, K.; Boxer, S. G. *Science (80-.).* **1994**, *264* (5160), 810–816.

(38) Chattoraj, M.; King, B. a; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (16), 8362–8367.

(39) Bublitz, G.; King, B. A.; Boxer, S. G. *J. Am. Chem. Soc.* **1998**, *120* (36), 9370–9371.

(40) Oh, D. H.; Boxer, S. G. *J. Am. Chem. Soc.* **1990**, *112* (17), 8161–8163.

(41) Oh, D. H.; Sano, M.; Boxer, S. G. *J. Am. Chem. Soc.* **1991**, *113* (18), 6880–6890.

(42) Pierce, D. W.; Boxer, S. G. *Biophys. J.* **1995**, *68* (4), 1583–1591.

(43) Franzen, S.; Moore, L. J.; Woodruff, W. H.; Boxer, S. G. *J. Phys. Chem. B* **1999**, *103*, 3070–3072.

(44) Park, E. S.; Thomas, M. R.; Boxer, S. G. *J. Am. Chem. Soc.* **2000**, *122* (49), 12297–12303.

(45) Park, E. S.; Boxer, S. G. *J. Phys. Chem. B* **2002**, *106* (22), 5800–5806.

(46) Fried, S. D.; Bagchi, S.; Boxer, S. G. *J. Am. Chem. Soc.* **2013**, *135*, 11181–11192.

(47) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2000**, *104* (51), 11853–11863.

(48) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2002**, *106* (3), 469–477.

(49) Webb, L. J.; Boxer, S. G. *Biochemistry* **2008**, *47* (6), 1588–1598.

(50) Suydam, I. T.; Boxer, S. G. *Biochemistry* **2003**, *42* (41), 12050–12055.

(51) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science (80-.).* **2006**, *313* (July), 200–204.

(52) Russell, S. . T. .; Warshel, A. **1985**, *185*, 389–404.

(53) Gilson, M. K.; Honig, B. H. *Nature* **1987**, *330* (6143), 84–86.

(54) Stafford, A. J.; Walker, D. M.; Webb, L. J. *Biochemistry* **2012**, *51* (13), 2757–2767.

(55) Fafarman, A. T.; Boxer, S. G. *J. Phys. Chem. B* **2010**, *114* (42), 13536–13544.

(56) Fafarman, A. T.; Sigala, P. A.; Schwans, J. P.; Fenn, T. D.; Herschlag, D.; Boxer, S. G. *Proc. Natl. Acad. Sci.* **2012**, *109* (6), E299–E308.

(57) Fried, S. D.; Bagchi, S.; Boxer, S. G. *Science* **2014**, *346* (6216), 1510–1514.

(58) Wang, L.; Brock, a; Herberich, B.; Schultz, P. G. *Science* **2001**, *292* (5516), 498–500.

(59) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128* (43), 13984–13985.

(60) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. *Nat. Protoc.* **2007**, *2* (10), 2601–2607.

(61) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. *J. Am. Chem. Soc.* **2006**, *128* (41), 13356–13357.

(62) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

(63) Bagchi, S.; Fried, S. D.; Boxer, S. G. *J. Am. Chem. Soc.* **2012**, *134*, 10373–10376.

(64) Stafford, A. J.; Ensign, D. L.; Webb, L. J. *J. Phys. Chem. B* **2010**, *114* (46), 15331–15344.

(65) Walker, D. M.; Wang, R.; Webb, L. J. *Phys. Chem. Chem. Phys.* **2014**, *16* (37), 20047–20060.

(66) Ritchie, A. W.; Webb, L. J. *J. Phys. Chem. B* **2014**, *118* (28), 7692–7702.

(67) Ritchie, A. W.; Webb, L. J. *J. Phys. Chem. B* **2015**, *119* (44), 13945–13957.

(68) Reimers, J. R.; Hall, L. E. *J. Am. Chem. Soc.* **1999**, *121* (15), 3730–3744.

(69) Aschaffenburg, D. J.; Moog, R. S. *J. Phys. Chem. B* **2009**, *113* (38), 12736–12743.

(70) Getahun, Z.; Huang, C. Y.; Wang, T.; De León, B.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125* (2), 405–411.

(71) Choi, J.-H.; Oh, K.-I.; Lee, H.; Lee, C.; Cho, M. *J. Chem. Phys.* **2008**, *128* (13), 134506.

(72) Adhikary, R.; Zimmermann, J.; Dawson, P. E.; Romesberg, F. E. *ChemPhysChem* **2014**, 849–853.

(73) Fafarman, A. T.; Sigala, P. a.; Herschlag, D.; Boxer, S. G. *J. Am. Chem. Soc.* **2010**,

132 (37), 12811–12813.

(74) Zhang, W.; Markiewicz, B. N.; Doerksen, R. S.; Smith, III, A. B.; Gai, F. *Phys. Chem. Chem. Phys.* **2015**.

(75) Ragain, C. M.; Newberry, R. W.; Ritchie, A. W.; Webb, L. J. *J. Phys. Chem. B* **2012**, *116* (31), 9326–9336.

(76) Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (24), 13617–13622.

(77) Brejc, K.; Sixma, T. K.; Kitts, P. a; Kain, S. R.; Tsien, R. Y.; Ormo, M.; Remington, S. J. *Proc. Natl. Acad. Sci.* **1997**, *94* (6), 2306–2311.

(78) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. *Science* **1994**, *263* (5148), 802–805.

(79) Ormö, M.; Cubitt, a B.; Kallio, K.; Gross, L. a; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273* (5280), 1392–1395.

(80) Meech, S. R. *Chem. Soc. Rev.* **2009**, *38* (10), 2922–2934.

(81) Wachter, R. M.; Elsliger, M.; Kallio, K.; Hanson, G. T.; Remington, S. J. *Structure* **1998**, *6* (10), 1267–1277.

(82) Patterson, G. H.; Lippincott-Schwartz, J. *Science (80-.).* **2002**, *297* (September), 1873–1877.

(83) Heim, R.; Cubitt, a B.; Tsien, R. Y. *Nature* **1995**, *373* (6516), 663–664.

(84) Jung, G.; Wiehler, J.; Zumbusch, A. *Biophys. J.* **2005**, *88* (3), 1932–1947.

(85) van Thor, J. J.; Gensch, T.; Hellingwerf, K. J.; Johnson, L. N. *Nat. Struct. Biol.* **2002**, *9* (1), 37–41.

(86) Nienhaus, K.; Nienhaus, G. U. *Chem Soc Rev* **2014**, *43* (4), 1088–1106.

(87) van Thor, J. J. *Chem. Soc. Rev.* **2009**, *38* (10), 2935–2950.

(88) Lukyanov, K. a; Chudakov, D. M.; Lukyanov, S.; Verkhusha, V. V. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 885–890.

(89) Lubbeck, J. L.; Dean, K. M.; Ma, H.; Palmer, A. E.; Jimenez, R. *Anal Chem* **2012**, *84* (9), 3929–3937.

(90) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat.*

Biotechnol. **2006**, *24* (1), 79–88.

(91) Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J. **1999**, 5296–5301.

(92) Wachter, R. M.; Elsliger, M. A.; Kallio, K.; Hanson, G. T.; Remington, S. J. *Structure* **1998**, *6* (10), 1267–1277.

(93) Bell, a F.; Stoner-Ma, D.; Wachter, R. M.; Tonge, P. J. *J. Am. Chem. Soc.* **2003**, *125* (5), 6919–6926.

(94) Henderson, J. N.; Gepshtein, R.; Heenan, J. R.; Kallio, K.; Huppert, D.; Remington, S. J. *J. Am. Chem. Soc.* **2009**, *131* (12), 4176–4177.

(95) Ding, L.; Chung, L. W.; Morokuma, K. *J. Phys. Chem. B* **2013**, *117* (4), 1075–1084.

(96) Patterson, G. H. *Proc. SPIE* **2004**, *5329*, 13–22.

(97) Betzig, E.; Patterson, G. H.; Sougrat, R.; Lindwasser, O. W.; Olenych, S.; Bonifacino, J. S.; Davidson, M. W.; Lippincott-Schwartz, J.; Hess, H. F. *Science* **2006**, *313* (5793), 1642–1645.

(98) Berlin, S.; Carroll, E. C.; Newman, Z. L.; Okada, H. O.; Quinn, C. M.; Kallman, B.; Rockwell, N. C.; Martin, S. S.; Lagarias, J. C.; Isacoff, E. Y. *Nat. Methods* **2015**, *12* (9), 852–858.

(99) Subach, F. V; Malashkevich, V. N.; Zencheck, W. D.; Xiao, H.; Filonov, G. S.; Almo, S. C.; Verkhusha, V. V. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (50), 21097–21102.

(100) Fron, E.; De Keersmaecker, H.; Rocha, S.; Baeten, Y.; Lu, G.; Uji-I, H.; Van Der Auweraer, M.; Hofkens, J.; Mizuno, H. *J. Phys. Chem. B* **2015**, *119* (47), 14880–14891.

(101) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35* (1), 225–249.

(102) Lang, K.; Chin, J. W. **2014**.

(103) Young, T. S.; Schultz, P. G. *J. Biol. Chem.* **2010**, *285* (15), 11039–11044.

(104) Miyake-Stoner, S. J.; Refakis, C. A.; Hammill, J. T.; Lusic, H.; Hazen, J. L.; Deiters, A.; Mehl, R. A. *Biochemistry* **2010**, *49* (8), 1667–1677.

(105) Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8* (9), 883–890.

(106) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79* (1), 413–444.

(107) Dippel, A. B.; Olenginski, G. M.; Maurici, N.; Liskov, M. T.; Brewer, S. H.; Phillips-Piro, C. M. *Acta Crystallogr. Sect. D Struct. Biol.* **2016**, *72* (1), 1–10.

(108) Rosell, F. I.; Boxer, S. G. *Biochemistry* **2003**, *42* (1), 177–183.

(109) Henderson, J. N.; Gepshtein, R.; Heenan, J. R.; Kallio, K.; Huppert, D.; Remington, S. J. *J. Am. Chem. Soc.* **2009**, *131* (12), 4176–4177.

(110) Purcell, K. F.; Drago, R. S. *J. Am. Chem. Soc.* **1966**, *88* (5), 919–924.

(111) Fawcett, W. R.; Liu, G.; Kessler, T. E. *J. Phys. Chem.* **1993**, 9293–9298.

(112) Zimmermann, J.; Thielges, M. C.; Seo, Y. J.; Dawson, P. E.; Romesberg, F. E. *Angew. Chemie - Int. Ed.* **2011**, *50*, 8333–8337.

(113) Adhikary, R.; Zimmermann, J.; Dawson, P. E.; Romesberg, F. E. *Anal. Chem.* **2015**, acs.analchem.5b03437.

(114) Slocum, J. D.; Webb, L. J. *J. Am. Chem. Soc.* **2016**, *138* (20), 6561–6570.

(115) The Mathworks, Inc. 2016.

(116) Stoner-Ma, D.; Jaye, A. A.; Matousek, P.; Towrie, M.; Meech, S. R.; Tonge, P. J. *J. Am. Chem. Soc.* **2005**, *127* (9), 2864–2865.

(117) Bogdanov, A. M.; Acharya, A.; Titelmayer, A. V.; Mamontova, A. V.; Bravaya, K. B.; Kolomeisky, A. B.; Lukyanov, K. A.; Krylov, A. I. *J. Am. Chem. Soc.* **2016**, *138* (14), 4807–4817.

(118) Grigorenko, B. L.; Nemukhin, A. V.; Morozov, D. I.; Polyakov, I. V.; Bravaya, K. B.; Krylov, A. I. *J. Chem. Theory Comput.* **2012**, *8* (6), 1912–1920.

(119) Maienschein-Cline, M. G.; Londergan, C. H. *J. Phys. Chem. A* **2007**, *111* (40), 10020–10025.

(120) Haldar, T.; Kashid, S. M.; Deb, P.; Kesh, S.; Bagchi, S. *J. Phys. Chem. Lett.* **2016**, *7* (13), 2456–2460.

(121) Deb, P.; Haldar, T.; Kashid, S. M.; Banerjee, S.; Chakrabarty, S.; Bagchi, S. *J. Phys. Chem. B* **2016**, *120* (17), 4034–4046.

(122) Shaner, N. C.; Patterson, G. H.; Davidson, M. W. *J. Cell Sci.* **2007**, *120*, 4247–

4260.